



*Discussion paper 2014-01*

# **INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS**

**Marc Henry  
Koen Jochmans  
Bernard Salanié**

*Sciences Po Economics Discussion Papers*

## INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS\*

KOEN JOCHMANS<sup>1</sup> and MARC HENRY<sup>2</sup> and BERNARD SALANIÉ<sup>3</sup>

<sup>1</sup>*Sciences Po, 28 rue des Saint Pères, 75007 Paris, France.*

*E-mail:* [koen.jochmans@sciencespo.fr](mailto:koen.jochmans@sciencespo.fr)

<sup>2</sup>*The Pennsylvania State University, University Park, PA 16801, U.S.A.*

*E-mail:* [marc.henry@psu.edu](mailto:marc.henry@psu.edu)

<sup>3</sup>*Columbia University, 420 West 118th Street, New York, NY 10027, U.S.A.*

*E-mail:* [bsalanie@columbia.edu](mailto:bsalanie@columbia.edu)

Revised on March 27, 2015

Many econometric models can be analyzed as finite mixtures. We focus on two-component mixtures and we show that they are nonparametrically point identified by a combination of an exclusion restriction and tail restrictions. Our identification analysis suggests simple closed-form estimators of the component distributions and mixing proportions. We derive their asymptotic properties using results on tail empirical processes and we present a simulation study that documents their finite-sample performance.

*Keywords:* mixture model, nonparametric identification and estimation, tail empirical process.

## INTRODUCTION

The use of finite mixtures has a long history in applied econometrics. A non-exhaustive list of applications includes models with discrete unobserved heterogeneity, hidden Markov chains, and models with mismeasured discrete variables; see [Henry et al. \(2014\)](#) for a more extensive discussion of applications. Until recently, the literature on nonparametric identification of mixture models was sparse. Following the lead of [Hall and Zhou \(2003\)](#), several authors have analyzed multivariate mixtures; recent contributions are [Kasahara and Shimotsu \(2009\)](#), [Allman et al. \(2009\)](#), and [Bonhomme et al. \(2014a,b\)](#). There are fewer identifying restrictions available when the model of interest is univariate. [Bordes](#)

---

\*We are grateful to Arthur Lewbel and three referees for comments and suggestions, and to Victor Chernozhukov and Yuichi Kitamura for fruitful discussions. Parts of this paper were written while Henry was visiting the University of Tokyo Graduate School of Economics and while Salanié was visiting the Toulouse School of Economics. The hospitality of both institutions is gratefully acknowledged. Henry's research has received funding from SSHRC Grants 410-2010-242 and 435-2013-0292, and NSERC Grant 356491-2013. Jochmans' research has received funding from the SAB grant "Nonparametric estimation of finite mixtures". Some of the results presented here previously circulated as part of [Henry et al. \(2010\)](#), whose published version ([Henry et al. 2014](#)) only contains results on partial identification.

et al. (2006), for instance, provide such restrictions for location models with symmetric error distributions.

In this paper we give sufficient conditions that point-identify univariate component distributions and associated mixing proportions. The restrictions we rely on are most effective in two-component models. To simplify the analysis we focus on this case, like Hall and Zhou (2003) and Bordes et al. (2006). We comment briefly on mixtures with more components at the end of the paper. Our arguments are constructive, and we propose closed-form estimators for both the component distributions and the mixing proportions. We derive their large-sample properties and investigate their behavior in a simulation experiment.

The model we consider in this paper is characterized by an exclusion restriction and a tail-dominance assumption. Like Henry et al. (2014), we assume the existence of a source of variation that shifts the mixing proportions but leaves the component distributions unchanged. Such an assumption is natural in several important applications, such as measurement-error models (Mahajan 2006). In hidden Markov models, it follows directly from the model specification. The exclusion restriction is also implied by the conditional-independence restriction that underlies the results of Hall and Zhou (2003) and others on multivariate mixtures.

It was shown in Henry et al. (2014) that our exclusion restriction restricts both the component distributions and the mixing proportions to lie in a non-trivial set. Here we achieve point-identification by complementing the exclusion restriction with a restriction on the relative tail behavior of the component distributions. This restriction is quite natural in location models, for instance, but it can be motivated more generally. For example, regime-switching models typically feature regimes with different tail behavior. Alternatively, theoretical models can imply the required tail behavior; an example is the search and matching model of Shimer and Smith (2000), as explained in D'Haultfoeuille and Février (2014).

Our identification argument suggests plug-in estimators of the mixing proportions and the component distributions that are available in closed form. The estimators are based on ratios of intermediate quantiles, and their convergence rate is determined by the theory of tail empirical processes. As we rely on the tail behavior of the component

distributions to infer the mixing proportions, our estimators converge more slowly than the parametric rate. If the mixing proportions were known—or could be estimated at the parametric rate—the tail restrictions could be dispensed with and the implied estimator of the component distributions would also converge at the parametric rate.

Our estimators are consistent under very weak tail dominance assumptions. To control for asymptotic bias in their limit distribution we need to impose stronger requirements that prevent the tails of the components from vanishing too quickly. These assumptions rule out the Gaussian location model. Such thin-tailed distributions are known to be problematic for inference techniques that rely on tail behavior ([Khan and Tamer 2010](#)). However, we show that our assumptions apply to distributions with fatter tails, such as Pareto distributions.

Identification only requires that the variable subject to the exclusion restriction can take on two values. If it can take on more values, the model is overidentified and the specification can be tested. We derive the asymptotic properties of two statistics that can be used to respectively test the validity of the tail restrictions and that of the exclusion restriction.

The tail conditions we use to obtain nonparametric identification are related to the well-known identification-at-infinity argument of [Heckman \(1990\)](#); see also [D’Haultfoeulle and Maurel \(2013\)](#) for another approach. Other types of support restrictions have been used in related problems to establish identification. [Schwarz and Van Bellegem \(2010\)](#) imposed support restrictions in a semiparametric deconvolution problem to deal with measurement error in location models. [D’Haultfoeulle and Février \(2014\)](#) relied on a support condition as an alternative to completeness conditions ([Hu and Schennach 2008](#)) in multivariate mixture models.

The remainder of the paper is organized as follows. Section 1 describes the mixture model and proves identification. We rely on these results to construct estimators and derive their asymptotic properties in Section 2. Section 3 describes our specification tests. In Section 4 we conduct a Monte Carlo experiment that gives evidence on the small-sample performance of our methods. Finally, we conclude with some remarks on mixtures with more than two components.

## 1. MIXTURES WITH EXCLUSION AND TAIL RESTRICTIONS

Let  $(Y, X) \in \mathbb{R} \times \mathcal{X}$  be random variables. We assume throughout that our mixtures satisfy the following simple exclusion restriction.<sup>1</sup>

ASSUMPTION 1 (Mixture with exclusion).  *$F(y|x) \equiv \mathbb{P}(Y \leq y|X = x)$  decomposes as the two-component mixture*

$$F(y|x) = G(y) \lambda(x) + H(y) (1 - \lambda(x)) \quad (1.1)$$

*for distribution functions  $G : \mathbb{R} \mapsto [0, 1]$  and  $H : \mathbb{R} \mapsto [0, 1]$  and a function  $\lambda : \mathcal{X} \mapsto [0, 1]$  that maps values  $x$  into mixing proportions.*

The assumption that the component distributions do not depend on  $X$  embodies our exclusion restriction; see also Henry et al. (2014).

We complete the mixture model with the following assumption.

ASSUMPTION 2. *The mixing proportion  $\lambda$  is non-constant on  $\mathcal{X}$  and is bounded away from zero and one on  $\mathcal{X}$ .*

Non-constancy of  $\lambda$  gives the variable  $X$  relevance. Bounding  $\lambda$  away from zero and one implies that the mixture is irreducible.<sup>2</sup>

### 1.1. Motivating examples

Our first example has a long history in empirical work (Frisch 1934).

EXAMPLE 1 (Mismeasured treatments). *Let  $T$  denote a binary treatment indicator. Suppose that  $T$  is subject to classification error: rather than observing  $T$ , we observe misclassified treatment  $X$ . The distribution of the outcome variable  $Y$  given  $X = x$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1, X = x) \lambda(x) + \mathbb{P}(Y \leq y|T = 0, X = x) (1 - \lambda(x)),$$

---

<sup>1</sup>We omit conditioning variables throughout. The identification analysis extends straightforwardly. In principle, the distribution theory could be extended by using local empirical process results along the lines of Einmahl and Mason (1997). We postpone a detailed investigation into such an extension to future work.

<sup>2</sup>Note that irreducibility rules out the possibility of achieving identification of  $G$  and  $H$  via an identification-at-infinity argument, as in Heckman (1990) and Andrews and Schafgans (1998) for instance.

with  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . The usual ignorability assumption states that  $X$  and  $Y$  are independent given  $T$ . That is,

$$\mathbb{P}(Y \leq y|T = t, X = x) = \mathbb{P}(Y \leq y|T = t),$$

for  $t \in \{0, 1\}$ , in which case the decomposition of  $F(y|x)$  reduces to the model in (1.1) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$  and  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ . Also note that  $\lambda$  is non-constant unless misclassification in  $T$  is completely random.

The identification of treatment effects when the treatment indicator is mismeasured has received considerable attention, especially in the context of regression models (Bollinger 1996; Mahajan 2006; Lewbel 2007). Here, the conditional ignorability assumption that validates our exclusion restriction relies on non-differential misclassification error. It has been routinely used elsewhere (Carroll et al. 2006).

Our second example deals with regime-switching models, also referred to as hidden Markov models. These models cover switching regressions, which have been used in a variety of settings (see, e.g., Heckman 1974, Hamilton 1989), as well as several versions of stochastic-volatility models (Ghysels et al. 1996).

EXAMPLE 2 (Hidden Markov model). *Let  $\mathbf{Y} = (Y_1, \dots, Y_T)'$  be a time series of outcome variables. A hidden Markov model for the dependency structure in these data assumes that there is a discrete latent series of state variables  $\mathbf{S} = (S_1, \dots, S_T)'$  having Markov dependence, that the variables in  $\mathbf{Y}$  are jointly independent given  $\mathbf{S}$ , and that*

$$\mathbb{P}(Y_t \leq y_t|\mathbf{S} = \mathbf{s}) = \mathbb{P}(Y_t \leq y_t|S_t = s_t).$$

*To see that such a model fits (1.1), assume that there are two latent states 0 and 1 and (for notational simplicity) that  $\mathbf{S}$  has first-order Markov dependence. Denote  $\mathbf{X} = (Y_1, \dots, Y_{t-1})'$ . Then*

$$F(y_t|\mathbf{x}) = \mathbb{P}(Y_t \leq y_t|S_t = 1)\mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x}) + \mathbb{P}(Y_t \leq y_t|S_t = 0)\mathbb{P}(S_t = 0|\mathbf{X} = \mathbf{x}),$$

*which fits our setup. Moreover,  $\lambda(\mathbf{x}) = \mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x})$  does vary with  $\mathbf{x}$ , unless the outcomes are independent of the latent states.*

In this example, the exclusion restriction follows directly from the Markovian structure of the regime-switching model. [Gassiat and Rousseau \(2014\)](#) obtained nonparametric identification in location models when the matrix of transition probabilities of the Markov chain has full rank. The approach presented here delivers nonparametric identification in a much broader range of models.

Our third example links [\(1.1\)](#) to the recent literature on multivariate mixtures building on [Hall and Zhou \(2003\)](#).

**EXAMPLE 3 (Multivariate mixture).** *Suppose  $Y$  and  $X$  are two measurements that are independent conditional on a latent binary factor  $T$ :*

$$\begin{aligned}\mathbb{P}(Y \leq y, X \leq x) &= \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(X \leq x|T = 1) \mathbb{P}(T = 1) \\ &\quad + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(X \leq x|T = 0) \mathbb{P}(T = 0).\end{aligned}$$

*Then the conditional distribution of the  $Y$  given  $X$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(T = 1|X = x) + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(T = 0|X = x).$$

*This is of the form in [\(1.1\)](#) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$ ,  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ , and  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . Note that the bivariate mixture model implies that the distribution of  $X$  given  $Y$  decomposes in the same way.*

[Hall and Zhou \(2003\)](#) showed that multivariate two-component mixtures with conditional independence restrictions are nonparametrically identified from data on three or more measurements and are set identified from data on only two measurements. The results we derive below imply that two measurements can also yield point identification under tail restrictions.

## 1.2. Identification

We show below that both the mixture components  $G, H$  and the mixing proportions  $\lambda$  are identified under the following dominance condition on the tails of the component distributions.

ASSUMPTION 3 (Tail dominance).

(i) The left tail of  $G$  is thinner than the left tail of  $H$ , i.e.,

$$\lim_{y \downarrow -\infty} \frac{G(y)}{H(y)} = 0.$$

(ii) The right tail of  $G$  is thicker than the right tail of  $H$ , i.e.,

$$\lim_{y \uparrow +\infty} \frac{1 - H(y)}{1 - G(y)} = 0.$$

Tail dominance is natural in location models.

EXAMPLE 4 (Location models). Suppose that  $Y = \mu(T) + U$ , where  $T$  is a binary indicator and  $U \sim F$ , independent of  $T$ . Then (1.1) yields

$$F(y|x) = F(y - \mu(1)) \mathbb{P}(T = 1|X = x) + F(y - \mu(0)) \mathbb{P}(T = 0|X = x).$$

Suppose that  $\mu(0) < \mu(1)$ , that  $F$  is absolutely continuous, and that the derivatives of  $-\ln(1 - F(u))$  and  $-\ln F(u)$  grow without bound as  $u \uparrow +\infty$  and  $u \downarrow -\infty$ , respectively. Then Assumption 3 holds with  $G(y) = F(y - \mu(1))$  and  $H(y) = F(y - \mu(0))$ .

**Proof.** We prove Assumption 3(ii). Let  $\varphi(u) \equiv -\ln(1 - F(u))$  and denote  $\varphi'$  its first derivative. Then

$$\frac{1 - F(y - \mu(0))}{1 - F(y - \mu(1))} = \exp(\varphi(y - \mu(1)) - \varphi(y - \mu(0))) = \exp(-\varphi'(y^*)(\mu(1) - \mu(0))),$$

where  $y^*$  lies between  $y - \mu(1)$  and  $y - \mu(0)$ . Since  $\varphi'$  increases without bound as  $y \uparrow +\infty$  and  $\mu(1) > \mu(0)$ , the expression on the right-hand side tends to zero as  $y$  increases. Assumption 3(i) can be verified in the same way.  $\square$

We now show that, combined, our exclusion restriction and tail dominance identify all elements of the mixture model.

THEOREM 1 (Identification). Under Assumptions 1–3,  $G$ ,  $H$ , and  $\lambda$  are identified.



**Proof.** The proof is constructive. Fix  $x' \in \mathcal{X}$  and choose  $x'' \in \mathcal{X}$  so that  $\lambda(x') \neq \lambda(x'')$ . Then re-arranging (1.1) gives

$$\begin{aligned} \frac{F(y|x')}{F(y|x'')} &= \frac{1 + \lambda(x') (G(y)/H(y) - 1)}{1 + \lambda(x'') (G(y)/H(y) - 1)}, \\ \frac{1 - F(y|x')}{1 - F(y|x'')} &= \frac{\lambda(x') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x'))}{\lambda(x'') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x''))}. \end{aligned}$$

Taking limits, Assumption 3 further implies that

$$\begin{aligned} \zeta^-(x', x'') &\equiv \lim_{y \downarrow -\infty} \frac{F(y|x')}{F(y|x'')} = \frac{1 - \lambda(x')}{1 - \lambda(x'')}, \\ \zeta^+(x', x'') &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|x')}{1 - F(y|x'')} = \frac{\lambda(x')}{\lambda(x'')}. \end{aligned} \tag{1.2}$$

These two equations can be solved for the mixing proportion at  $x'$ , yielding

$$\lambda(x') = \frac{1 - \zeta^-(x'', x')}{\zeta^+(x'', x') - \zeta^-(x'', x')}. \tag{1.3}$$

Since  $\lambda$  is non-constant, for any  $x' \in \mathcal{X}$  there exists a  $x'' \in \mathcal{X}$  for which such a system of equations can be constructed. The function  $\lambda$  is therefore identified on its entire support.

To establish identification of  $G$  and  $H$ , first note that

$$G(y) - H(y) = \frac{F(y|x'') - F(y|x')}{\lambda(x'') - \lambda(x')} \tag{1.4}$$

follows from (1.1). Then, evaluating (1.1) in  $x''$  and re-arranging the resulting expression for  $F(y|x'')$  gives

$$H(y) = F(y|x'') - (G(y) - H(y)) \lambda(x'') = F(y|x'') - \frac{\lambda(x'')}{\lambda(x'') - \lambda(x')} (F(y|x'') - F(y|x')),$$

which is identified. Furthermore, using (1.2) we can write

$$H(y) = F(y|x'') - \frac{1}{1 - \zeta^+(x', x'')} (F(y|x'') - F(y|x')). \tag{1.5}$$

Plugging this expression for  $H(y)$  back into the mixture representation of  $F(y|x'')$  as in (1.1) further yields

$$G(y) = F(y|x'') - \frac{1}{1 - \zeta^-(x', x'')} (F(y|x'') - F(y|x')), \tag{1.6}$$

again using (1.2). This shows that both component distributions are identified, concluding the proof.  $\square$

If we only assume one-sided tail dominance, then either  $G$  or  $H$  remains identified.

**COROLLARY 1** (One-sided tail dominance). *Under Assumptions 1 and 2,  $G$  is identified if Assumption 3(i) holds and  $H$  is identified if Assumption 3(ii) holds.*

**Proof.** We consider identification of  $H$ . Let  $x', x''$  be as in the proof of Theorem 1. Under Assumption 3(ii) we can still determine  $\zeta^+(x', x'') = \lambda(x')/\lambda(x'')$ , from which we can learn the ratio  $1/(1 - \zeta^+(x'', x'))$ . Together with (1.5) this yields  $H$ . This concludes the proof of the corollary.  $\square$

The following example illustrates the usefulness of Corollary 1.

**EXAMPLE 5** (Stochastic volatility). *Consider a two-regime stochastic volatility model, which is a special case of Example 2. Assume that the outcome variable  $Y$  has mean zero and conditional variance*

$$T\sigma_G^2 + (1 - T)\sigma_H^2$$

*for positive constants  $\sigma_G^2$  and  $\sigma_H^2$ . Suppose that  $\sigma_G^2 > \sigma_H^2$ . Then  $G$  is the distribution associated with a regime that is characterized by relatively higher volatility. In this case, both tails of  $G$  dominate those of  $H$ . Hence, in Assumption 3, Condition (ii) holds but Condition (i) fails. Nevertheless, the distribution of the lower-volatility regime, namely  $H$ , remains identified.*

Our identification result suggests plug-in estimators of the mixing proportions and the component distributions. The proof of Theorem 1, and Equations (1.3) and (1.5)–(1.6) in particular, further show that our mixture model yields overidentifying restrictions as soon as the instrument can take on more than two values. We turn to estimation in the next section. Afterwards we discuss the construction of specification tests under overidentification.

## 2. ESTIMATION

To motivate the construction of our estimators, we first note that the structure of the model in (1.1) continues to hold when we aggregate across  $x$ . Extending our notation to

$$F(y|A) \equiv \mathbb{P}(Y \leq y|X \in A), \quad \lambda(A) \equiv \sum_{x \in A} \lambda(x) \mathbb{P}(X = x|X \in A),$$

for any  $A \subset \mathcal{X}$ , we have

$$F(y|A) = G(y) \lambda(A) + H(y) (1 - \lambda(A)), \quad (2.1)$$

which is of the same form as (1.1). Furthermore, the proof of Theorem 1 continues to go through for (2.1); replacing  $x'$  with  $A$  and  $x''$  with  $\mathcal{X} - A$  does not alter the argument.

We will assume from now on that  $X$  is discrete. As will become apparent, this only entails a loss of generality for the estimation of the function  $\lambda$ , as our estimator will only yield a discretized approximation to it. Extending our results to continuous  $X$  would complicate the exposition greatly and we feel that it would only distract from our main argument.

We will work under the following sampling condition.

ASSUMPTION 4.  $(Y_1, X_1), \dots, (Y_n, X_n)$  is a random sample on  $(Y, X)$ .

For each  $A \subset \mathcal{X}$ , let

$$F_n(y|A) \equiv n_A^{-1} \sum_{i=1}^n 1\{Y_i \leq y, X_i \in A\},$$

where  $n_A \equiv \sum_{i=1}^n 1\{X_i \in A\}$ .

For each pair of disjoint subsets  $A, B$  of  $\mathcal{X}$  we can generalize (1.2) to

$$\begin{aligned} \zeta^-(A, B) &\equiv \lim_{y \downarrow -\infty} \frac{F(y|A)}{F(y|B)} = \frac{1 - \lambda(A)}{1 - \lambda(B)}, \\ \zeta^+(A, B) &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|A)}{1 - F(y|B)} = \frac{\lambda(A)}{\lambda(B)}. \end{aligned} \quad (2.2)$$

For any subsample of size  $m$  and integers  $\iota_m$  and  $\kappa_m$ , let  $\ell_m$  and  $r_m$  denote the  $(\iota_m + 1)$ th and  $(m - \kappa_m)$ th order statistics of  $Y$  in this subsample. We estimate the quantities in (2.2) by

$$\zeta_n^-(A, B) \equiv \frac{F_n(\ell_{n_B}|A)}{F_n(\ell_{n_B}|B)}, \quad \zeta_n^+(A, B) \equiv \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)}, \quad (2.3)$$

respectively. In our asymptotic theory, we will choose  $\iota_{n_B}$  and  $\kappa_{n_B}$  so that  $\ell_{n_B} \downarrow -\infty$  and  $r_{n_B} \uparrow +\infty$  as  $n \uparrow +\infty$  at an appropriate rate.

Estimators of both the mixing proportions and the component distributions follow readily along the lines of the proof of Theorem 1; see below. Since their asymptotic distribution will be driven by the large-sample behavior of the estimators of the quantities in (2.3), we start by deriving the statistical properties of these estimators.

### 2.1. Asymptotic theory for intermediate quantiles

Throughout this section we fix disjoint sets  $A, B$  and consider the asymptotic behavior of the estimators in (2.3).

Consistency only requires the following rate conditions.

ASSUMPTION 5 (Order statistics).  $\iota_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  and  $\kappa_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  as  $n \uparrow +\infty$ .

THEOREM 2 (Consistency). *If Assumptions 1–5 hold,*

$$\zeta_n^-(A, B) \xrightarrow{P} \zeta^-(A, B), \quad \zeta_n^+(A, B) \xrightarrow{P} \zeta^+(A, B),$$

as  $n \uparrow +\infty$ .

**Proof.** We prove the theorem for  $\zeta_n^+$ ; the proof for  $\zeta_n^-$  follows in a similar fashion. Write

$$\zeta_n^+ - \zeta^+ = (\zeta_n^+ - \zeta^{\kappa_{n_B}}) + (\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.4)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . For the second right-hand side term in (2.4) we have

$$\zeta^{\kappa_{n_B}} - \zeta^+ = \left( \frac{\lambda(A) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(A))}{\lambda(B) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(B))} - \frac{\lambda(A)}{\lambda(B)} \right) = O_p\left(\frac{1-H(r_{n_B})}{1-G(r_{n_B})}\right) = o_p(1),$$

by Assumptions 3(ii) and 5. To deal with the first right-hand side term in (2.4), recall that

$$\zeta_n^+ - \zeta^{\kappa_{n_B}} = \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)} - \frac{1 - F(r_{n_B}|A)}{1 - F(r_{n_B}|B)}.$$

Letting  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$  we thus have that

$$\begin{aligned} \zeta_n^+ - \zeta^{\kappa_{n_B}} &= \frac{(1 - F(r_{n_B}|A))\mathbb{G}_n(r_{n_B}|B)/\sqrt{n_B} - (1 - F(r_{n_B}|B))\mathbb{G}_n(r_{n_B}|A)/\sqrt{n_A}}{(1 - F_n(r_{n_B}|B))(1 - F(r_{n_B}|B))} \\ &= \frac{\sqrt{n_B}}{\kappa_{n_B}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right) \\ &= O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right), \end{aligned}$$

where the second equality uses  $1 - F_n(r_{n_B}|B) = \kappa_{n_B}/n_B$  and the last one follows by the law of the iterated logarithm for empirical processes. Thus, from Assumption 5 it follows that  $|\zeta_n^+ - \zeta^{\kappa_{n_B}}| = o_p(1)$ . This completes the proof.  $\square$

Deriving the limit distribution requires some more care, and three more assumptions. We first impose the following regularity condition on the component distributions.

ASSUMPTION 6. *G and H are absolutely continuous on  $\mathbb{R}$ .*

This assumption is very weak. Note that we do not require the existence of any moments of the component distributions. Thus, our results apply to heavy-tailed distributions such as Cauchy and Pareto distributions.

We will also complement Assumption 5 with an additional rate condition.

ASSUMPTION 7 (Order statistics (cont'd.)).  *$\iota_{n_B}/n_B \downarrow 0$  and  $\kappa_{n_B}/n_B \downarrow 0$  as  $n \uparrow +\infty$ .*

Where Assumption 5 required the order statistics to grow to ensure consistency, this assumption bounds this growth rate so that appropriately scaled versions of  $\zeta_n^+$  and  $\zeta_n^-$  have a limit distribution.

Finally, we will use an additional condition on the relative tails of the component distributions.

ASSUMPTION 8 (Tail rates).

- (i)  $G(\ell_{n_B})/H(\ell_{n_B}) = o_p(1/\sqrt{\iota_{n_B}})$ ; and
- (ii)  $(1 - H(r_{n_B}))/ (1 - G(r_{n_B})) = o_p(1/\sqrt{\kappa_{n_B}})$ .

Assumption 8 rules out distributions whose tails vanish too quickly and ensures that the limit distributions are free of asymptotic bias. We comment on Assumption 8 after we derive the limit distributions of our estimators.

Let  $\rho_{A,B} \equiv \mathbb{P}(X \in B)/\mathbb{P}(X \in A)$ . Note that  $0 < \rho_{A,B} < +\infty$  because of random sampling. Introduce

$$\begin{aligned}\sigma_-^2(A, B) &\equiv \zeta^-(A, B)^2 + \rho_{A,B} \zeta^-(A, B), \\ \sigma_+^2(A, B) &\equiv \zeta^+(A, B)^2 + \rho_{A,B} \zeta^+(A, B),\end{aligned}$$

Theorem 2 provides the asymptotic properties of the estimators in (2.2) and is the main building block for our subsequent results.

**THEOREM 3** (Asymptotic normality). *If Assumptions 1–8 hold, then as  $n \uparrow +\infty$ ,*

$$\begin{aligned}\sqrt{\iota_{n_B}}(\zeta_n^-(A, B) - \zeta^-(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, B)), \\ \sqrt{\kappa_{n_B}}(\zeta_n^+(A, B) - \zeta^+(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, B));\end{aligned}$$

*and these two estimators are asymptotically independent.*

**Proof.** We focus on the limit behavior of  $\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+)$  here; the proof of the result for  $\sqrt{\iota_n}(\zeta_n^- - \zeta^-)$  follows similar lines.

As in the proof of Theorem 2, write

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+) = \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) + \sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.5)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . Assumption 8 implies that

$$\sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+) = \sqrt{\kappa_{n_B}} O_p\left(\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})}\right) = o_p(1).$$

Hence, the second right-hand side term in (2.5) is asymptotically negligible.

We now turn to the first term in (2.5). From the proof of Theorem 2 we have that

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) = \sqrt{\frac{n_B}{\kappa_{n_B}}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right),$$

where  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$ . Let  $\alpha_n(u) \equiv \sqrt{n}(\mathcal{U}_n(u) - u)$  for  $\mathcal{U}_n$  the empirical cumulative distribution of an i.i.d. sample of size  $n$  from a uniform distribution on  $[0, 1]$ . By Assumption 6,  $F(y|S)$  is continuous in  $y$  for all  $S \subset \mathcal{X}$ . Therefore

$$\mathbb{G}_n(y|A) = \alpha_{n_A}(1 - F(y|A)) \text{ and } \mathbb{G}_n(y|B) = \alpha_{n_B}(1 - F(y|B))$$

by an application of the probability integral transform. Hence, we may write

$$\begin{aligned} \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) &= \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \\ &\quad - \sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A}(1 - F(r_{n_B}|A)). \end{aligned} \quad (2.6)$$

We study the asymptotic behavior of each of the right-hand side terms in turn.

Start with the first right-hand side term in (2.6). From the definition of the order statistic  $r_{n_B}$ , we find by adding and subtracting  $F_n(r_{n_B}|B)$  that

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} \left( 1 + \frac{\sqrt{n_B}}{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) \right);$$

or, defining  $\varepsilon_n \equiv -\sqrt{n_B}/\kappa_{n_B} \mathbb{G}_n(r_{n_B}|B)$ ,

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Therefore we can write

$$\zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) = \sqrt{2} \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right). \quad (2.7)$$

By the law of the iterated logarithm together with Assumption 5,

$$\varepsilon_n = -\frac{\sqrt{n_B}}{\kappa_{n_B}} O_{a.s.} \left( \sqrt{\ln \ln n_B} \right) = O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right) = o_{a.s.}(1).$$

Hence,  $(1 - \varepsilon_n)/2$  converges almost surely to  $1/2$ , and  $(1 - \varepsilon_n)/2 \in (0, 1)$  for  $n$  large enough. We may then apply Theorem 2.1 in Einmahl (1992) to establish the convergence in distribution of  $\sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right)$  to a normal random variable with mean zero and variance  $1/2$ . This, together with Equation (2.7) and an application of Slutsky's theorem, then implies that

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \zeta^{\kappa_{n_B}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \xrightarrow{d} \zeta^+ Z_B^+, \quad (2.8)$$

where  $Z_B^+$  is a standard normal random variable.

Now turn to the second right-hand side term in (2.6). First observe that

$$1 - F(r_{n_B}|A) = \zeta^{\kappa_{n_B}} (1 - F(r_{n_B}|B)) = \zeta^{\kappa_{n_B}} \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Using  $\rho_{A,B} = \lim_{n \uparrow +\infty} n_B/n_A$ , this gives

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) = \sqrt{2\rho_{A,B} \zeta^+} \sqrt{\frac{n_A}{2\tilde{\kappa}_{n_A}}} \alpha_{n_A} \left( \frac{2\tilde{\kappa}_{n_A}}{n_A} \frac{1 - \varepsilon_n}{2} \right) + o_p(1),$$

where  $\tilde{\kappa}_{n_A} \equiv (\kappa_{n_B} \zeta^{\kappa_{n_B}})/(n_B/n_A)$ . As  $\tilde{\kappa}_{n_A}$  satisfies Assumption (1.5) of Theorem 2.1 in Einmahl (1992) we may apply his theorem again to obtain

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) \xrightarrow{d} \sqrt{\rho_{A,B} \zeta^+} Z_A^+, \quad (2.9)$$

where  $Z_A^+$  is a standard-normal random variable which, by random sampling, is independent of  $Z_B^+$ .

Combining (2.6) with (2.8) and (2.9) then gives

$$\sqrt{\kappa_{n_B}} (\zeta_n^+ - \zeta^{\kappa_{n_B}}) \xrightarrow{d} \zeta^+ Z_B^+ - \sqrt{\rho_{A,B} \zeta^+} Z_A^+,$$

as claimed. This concludes the proof.  $\square$

We finish this section with two examples that specialize Assumption 8 to densities with log-concave tails and Pareto tails, respectively. In both cases, Assumption 8 is implied by the rate conditions in Assumption 7.

EXAMPLE 6 (Log-concave tails). *Suppose that  $G$  and  $H$  have log-concave tails; and for notational simplicity, assume that*

$$-\ln(1 - G(y)) \sim \left( \frac{y}{\sigma_G^+} \right)^{\alpha_G^+}, \quad -\ln(1 - H(y)) \sim \left( \frac{y}{\sigma_H^+} \right)^{\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for real numbers  $\alpha_G^+, \alpha_H^+ > 1$  and  $\sigma_G^+, \sigma_H^+ > 0$ , and

$$-\ln G(y) \sim \left( \frac{-y}{\sigma_G^-} \right)^{\alpha_G^-}, \quad -\ln H(y) \sim \left( \frac{-y}{\sigma_H^-} \right)^{\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$



for real numbers  $\alpha_G^-, \alpha_H^- > 1$  and  $\sigma_G^-, \sigma_H^- > 0$ . Then Assumption 7 implies Assumption 8 if both

(i)  $\alpha_G^+ < \alpha_H^+$ , or  $\alpha_G^+ = \alpha_H^+$  and  $\sigma_G^+ > \sigma_H^+$ ; and

(ii)  $\alpha_G^- > \alpha_H^-$ , or  $\alpha_G^- = \alpha_H^-$  and  $\sigma_G^- < \sigma_H^-$

hold.

**Proof.** We verify the second rate; the first follows similarly. Throughout, fix the set  $B$ . Assumptions 3(ii) and 7 imply that

$$\begin{aligned} 1 - F(r_{n_B}|B) &= (1 - G(r_{n_B})) \lambda(B) + (1 - H(r_{n_B})) (1 - \lambda(B)) \\ &= (1 - G(r_{n_B})) (\lambda(B) + o_p(1)). \end{aligned}$$

Further, because  $\kappa_{n_B}/n_B = 1 - F_n(r_{n_B}|B)$ , adding and subtracting  $F(r_{n_B}|B)$  gives

$$\begin{aligned} \frac{\kappa_{n_B}}{n_B} &= (1 - F(r_{n_B}|B)) + (F_n(r_{n_B}|B) - F(r_{n_B}|B)) \\ &= (1 - F(r_{n_B}|B)) + O_{a.s.}(\sqrt{(\ln \ln n_B)/n_B}). \end{aligned}$$

Because  $(\ln \ln n_B)/n_B \rightarrow 0$ , put together, we find

$$\frac{\kappa_{n_B}}{n_B} = C (1 - G(r_{n_B})) (1 + o_p(1))$$

for some constant  $C$ . With  $G$  and  $H$  having log-concave tails, it then follows from this expression that, asymptotically,  $r_{n_B}$  behaves like  $\alpha_G^+ \sqrt[n_B]{\ln n_B}$ . So, because

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} \sim \exp \left\{ \left( \frac{r_{n_B}}{\sigma_G^+} \right)^{\alpha_G^+} - \left( \frac{r_{n_B}}{\sigma_H^+} \right)^{\alpha_H^+} \right\},$$

we have that

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} = \begin{cases} O_p \left( \exp(-(\ln n_B)^{\alpha_H^+/\alpha_G^+}) \right) & \text{if } \alpha_H^+ > \alpha_G^+ \\ O_p(1/n_B) & \text{if } \alpha_H^+ = \alpha_G^+ \text{ and } \sigma_H^+ < \sigma_G^+ \end{cases},$$

from which the conclusion follows.  $\square$

Example 6 does not cover location models with log-concave distributions for which the  $\alpha$  and  $\sigma$  parameters of  $H$  equal those of  $G$ . This includes the location model with Gaussian errors, for which  $\alpha = 2$  and  $\sigma$  is the common standard error. While our estimator remains

consistent in such cases, we do not know of general results on tail empirical processes that would yield the asymptotic distribution of the estimator in this knife-edge case. To assess the extent to which the failure of Assumption 8 may play a role for inference, our simulation experiments in Section 4 include a Gaussian location model.

EXAMPLE 7 (Pareto tails). *Let  $C$  denote a generic constant. Suppose that  $G$  and  $H$  have Pareto tails, i.e.,*

$$(1 - G(y)) \sim C y^{-\alpha_G^+}, \quad (1 - H(y)) \sim C y^{-\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

*for positive real numbers  $\alpha_H^+ > \alpha_G^+$  and*

$$G(y) \sim C (-y)^{-\alpha_G^-}, \quad H(y) \sim C (-y)^{-\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

*for positive real numbers  $\alpha_G^- < \alpha_H^-$ . Then Assumption 7 implies Assumption 8.*

**Proof.** The argument is very similar to the one that was used to verify Example 6. We treat the right tail only. We have

$$\frac{\kappa_{n_B}}{n_B} = (1 - G(r_{n_B})) (1 + o_p(1)) = C r^{-\alpha_G^+} (1 + o_p(1)).$$

Assumption 8 requires that  $(1 - H(r_{n_B})) / (1 - G(r_{n_B})) = o(1/\sqrt{\kappa_n})$ , that is, that  $r_{n_B}^{\alpha_G^+ - \alpha_H^+} = o_p(1/\sqrt{\kappa_{n_B}})$ . This rate condition is satisfied when

$$\left( \frac{n_B}{\kappa_{n_B}} \right)^{\frac{\alpha_G^+ - \alpha_H^+}{\alpha_G^+}} = o_p \left( \frac{1}{\sqrt{\kappa_{n_B}}} \right),$$

which can be achieved by setting  $\kappa_{n_B} = o(n_B^{\gamma^+})$  for

$$\gamma^+ \equiv \frac{\alpha_H^+ - \alpha_G^+}{\alpha_H^+ - \alpha_G^+/2}. \quad (2.10)$$

This condition is weaker than Assumption 7 and is therefore implied by it.  $\square$

Example 7 shows that our methods are well suited to deal with Pareto tails. Pareto tails show up in many economic applications. A time-honored example is income and wealth distributions (Atkinson et al. 2011), which are often modeled as a log-normal for

most quantiles, combined with a Pareto right tail. More generally, “power laws” have become a popular tool in finance, in studies of firm growth, and in urban economics (see [Gabaix 2009](#) for a recent survey, and [Acemoglu et al. 2012](#) for an application to business cycles.) Many recent models of monopolistic competition, as used in international trade for instance, also assume that productivities are Pareto-distributed ([Arkolakis et al. 2012](#)).

Identification only requires that the tail index of  $H$  be larger than that of  $G$ , that is,  $\alpha_H > \alpha_G$ . Denote  $c = \alpha_H/\alpha_G > 1$ . Equation (2.10) then gives a convergence rate arbitrarily close to  $n^{-\beta/2}$  for  $\beta = 2(c-1)/(2c-1)$ . For example, if  $c = 2$  then  $\beta = 2/3$  and our estimators will converge slightly more slowly than  $n^{-1/3}$ . However, as  $c$  increases,  $\beta$  rapidly converges to one, and our estimators will converge at close to the  $n^{-1/2}$  parametric rate.

## 2.2. Mixing proportions

Fix  $x \in \mathcal{X}$  and consider estimating  $\lambda(x)$ . Set  $A = \mathcal{X} - x$  and  $B = x$  in (2.2) and solve for  $\lambda(x)$  to get

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}.$$

The mixing proportion  $\lambda$  need not be a strictly monotonic function. Estimating  $\lambda(x)$  by an average of plug-in estimates of (1.3) could therefore be problematic, as the denominator in (1.3) can be zero or be arbitrarily close to it for some pairs of values  $(x', x'')$ .

We instead estimate the mixing proportion at  $X = x$  by a plug-in estimator based on (2.3), that is,

$$\lambda_n(x) \equiv \frac{1 - \zeta_n^-(A, x)}{\zeta_n^+(A, x) - \zeta_n^-(A, x)}.$$

This estimator uses observations with  $X_i \neq x$  in a way that immunizes it against small or zero denominators.

To present the asymptotic variance of this estimator we need to define

$$\begin{aligned} d^-(x) &\equiv \frac{1 - \zeta^+(A, x)}{(\zeta^+(A, x) - \zeta^-(A, x))^2}, \\ d^+(x) &\equiv \frac{\zeta^-(A, x) - 1}{(\zeta^+(A, x) - \zeta^-(A, x))^2}. \end{aligned} \tag{2.11}$$

The speed of convergence and the asymptotic distribution of the  $\lambda_n(x)$  depend on the ratio  $c_x \equiv \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

**THEOREM 4** (Mixing proportions). *Under the conditions of Theorem 2,*

$$|\lambda_n(x) - \lambda(x)| = o_p(1)$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, d^-(x)^2 \sigma_-^2(A, x) + c_x d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x < +\infty, \\ \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, c_x^{-1} d^-(x)^2 \sigma_-^2(A, x) + d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x > 0, \end{aligned}$$

as  $n \uparrow +\infty$ .

**Proof.** The consistency claim follows directly from Theorem 2 by an application of the continuous mapping theorem.

To establish the asymptotic distribution, note that Theorem 2 states that

$$\begin{aligned} \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, x)), \\ \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, x)), \end{aligned}$$

and that  $\zeta_n^-(x)$  and  $\zeta_n^+(x)$  are asymptotically independent. An expansion around  $\zeta^-(A, x)$  and  $\zeta^+(A, x)$  then yields

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) \\ &\quad + d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) \sqrt{\frac{\iota_{n_x}}{\kappa_{n_x}}} + o_p(1), \end{aligned}$$

which has the limit distribution stated in the theorem if  $c_x$  is finite. Also, by the same argument,

$$\begin{aligned} \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(x) - \zeta^+(x)) \\ &\quad + d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(x) - \zeta^-(x)) \sqrt{\frac{\kappa_{n_x}}{\iota_{n_x}}} + o_p(1) \end{aligned}$$

converges in distribution as stated in the theorem if  $c_x$  is non-zero. This verifies the claims and proves the theorem.  $\square$

### 2.3. Component distributions

To estimate the component distributions let  $A, B$  be two sets that partition  $\mathcal{X}$ . Equations (1.5) and (1.6) then suggest the estimators

$$\begin{aligned} H_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^+(B, A)} (F_n(y|A) - F_n(y|B)), \\ G_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)). \end{aligned} \quad (2.12)$$

Fix  $A, B$  and omit them as arguments, so  $G_n(y) = G_n(y; A, B)$  and  $H_n(y) = H_n(y; A, B)$ .

To state their asymptotic behavior, let

$$\begin{aligned} d_G(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^-(B, A))^2}, \\ d_H(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^+(B, A))^2}, \end{aligned}$$

and let  $\|\cdot\|_\infty$  denote the supremum norm.

**THEOREM 5.** *Under the conditions of Theorem 2,*

$$\|G_n - G\|_\infty = o_p(1), \quad \|H_n - H\|_\infty = o_p(1),$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_A}}(G_n(y) - G(y)) &\xrightarrow{d} \mathcal{N}(0, d_G(A, B; y)^2 \sigma_-^2(B, A)), \\ \sqrt{\kappa_{n_A}}(H_n(y) - H(y)) &\xrightarrow{d} \mathcal{N}(0, d_H(A, B; y)^2 \sigma_+^2(B, A)), \end{aligned}$$

as  $n \uparrow +\infty$  for each  $y \in \mathbb{R}_+$ .

**Proof.** Consistency follows by Theorem 2 and the Glivenko-Cantelli theorem.

We establish the asymptotic distribution of  $G_n$ ; the result for  $H_n$  follows by the same argument.

First note that

$$\sqrt{\iota_{n_A}}(G_n(y) - G(y)) = T_1 + T_2 + T_3$$

for

$$\begin{aligned} T_1 &\equiv \sqrt{\iota_{n_A}}(F_n(y|A) - F(y|A)), \\ T_2 &\equiv -\frac{1}{1 - \zeta^-(B, A)} \sqrt{\iota_{n_A}} \left( \{F_n(y|A) - F(y|A)\} - \{F_n(y|B) - F(y|B)\} \right), \\ T_3 &\equiv -(F_n(y|A) - F_n(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right). \end{aligned}$$

By the Glivenko-Cantelli theorem,  $T_1 = o_p(1)$  and  $T_2 = o_p(1)$  while

$$T_3 = -(F(y|A) - F(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right) + o_p(1).$$

A linearization of this expression in  $\zeta_n^-(B, A) - \zeta^-(B, A)$  together with an application of Theorem 2 to the partition  $A, B$  then yields the result.  $\square$

### 3. SPECIFICATION TESTING

The tail restriction in Assumption 3 and the exclusion restriction on  $X$  are the two key ingredients that lead to identification via Theorem 1. The proof of the theorem shows that both restrictions have testable implications.

#### 3.1. Testing for tail dominance

As is evident from the proof of Theorem 1, Assumption 3 implies that

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}$$

for any  $A \subset \mathcal{X} - x$ . Denote  $\lambda_n(x; A)$  the corresponding plug-in estimator of  $\lambda(x)$ . Any such estimator will be consistent and asymptotically-normal under the conditions of Theorem 2 and Theorem 3, respectively. Therefore, letting  $A, B$  be two disjoint subsets of  $\mathcal{X} - x$ , a simple specification test for the tail restriction amounts to testing whether the difference

$$\lambda_n(x; A) - \lambda_n(x; B)$$

is statistically different from zero.

As before, the limit distribution of this difference depends on  $c_x = \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

We define

$$\begin{aligned} \Sigma_- &\equiv d^-(x, A) \left\{ d^-(x, A) \sigma_-^2(A, x) - d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) \right\} \\ &\quad + d^-(x, B) \left\{ d^-(x, B) \sigma_-^2(B, x) - d^-(x, A) \zeta^-(A, x) \zeta^-(B, x) \right\}, \end{aligned}$$

and

$$\begin{aligned} \Sigma_+ \equiv & d^+(x, A) \left\{ d^+(x, A) \sigma_+^2(A, x) - d^+(x, B) \zeta^+(A, x) \zeta^+(B, x) \right\} \\ & + d^+(x, B) \left\{ d^+(x, B) \sigma_+^2(B, x) - d^+(x, A) \zeta^+(A, x) \zeta^+(B, x) \right\}, \end{aligned}$$

in Theorem 6.

THEOREM 6 (A test for tail dominance). *Under the conditions of Theorem 3,*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{\Sigma_- + c_x \Sigma_+ / \sqrt{\ell_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x < +\infty$ , while

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{c_x^{-1} \Sigma_- + \Sigma_+ / \sqrt{\kappa_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x > 0$ , where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** Fix the sets  $A, B$ . We treat only the case  $c_x > 0$ . Our assumptions imply the joint limit result

$$\sqrt{\kappa_{n_x}} \begin{pmatrix} \lambda_n(x; A) - \lambda(x) \\ \lambda_n(x; B) - \lambda(x) \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\kappa^{AA} & \Sigma_\kappa^{AB} \\ \Sigma_\kappa^{AB} & \Sigma_\kappa^{BB} \end{pmatrix} \right),$$

where

$$\begin{aligned} \Sigma_\kappa^{AA} &\equiv c_x^{-1} d^-(x, A)^2 \sigma_-^2(A, x) + d^+(x, A)^2 \sigma_+^2(A, x), \\ \Sigma_\kappa^{BB} &\equiv c_x^{-1} d^-(x, B)^2 \sigma_-^2(B, x) + d^+(x, B)^2 \sigma_+^2(B, x), \end{aligned}$$

follow directly from an application of Theorem 4, and

$$\Sigma_\kappa^{AB} \equiv c_x^{-1} d^-(x, A) d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) + d^+(x, A) d^+(x, B) \zeta^+(A, x) \zeta^+(B, x)$$

follows readily from the asymptotic-linearity statement at the end of the proof of Theorem 3. It follows that the transformation  $\sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda_n(x))$  is asymptotically normal with variance

$$\Sigma_\kappa^{AA} + \Sigma_\kappa^{BB} - 2 \Sigma_\kappa^{AB} = c_x^{-1} \Sigma_- + \Sigma_+.$$

This proves the result. □

### 3.2. Testing the exclusion restriction

An implication of the exclusion restriction in (1.1) is that the estimators of the component distributions  $G$  and  $H$  in (2.12), when based on different subsets of  $\mathcal{X}$ , should co-incide with one another, up to sampling error.

Theorem 7 provides the relevant asymptotic variance to perform this test. In it we use

$$\begin{aligned}\Sigma_G = & d_G(A, C; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\} \\ & + d_G(A, B; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\}\end{aligned}$$

and

$$\begin{aligned}\Sigma_H = & d_H(A, C; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\} \\ & + d_H(A, B; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\},\end{aligned}$$

where the triple  $A, B, C$  constitutes any partition of  $\mathcal{X}$ .

**THEOREM 7** (A test of the exclusion restriction). *Under the conditions of Theorem 3*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{G_n(y; A, B) - G_n(y; A, C)}{\sqrt{\Sigma_G}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

and

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{H_n(y; A, B) - H_n(y; A, C)}{\sqrt{\Sigma_H}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** We consider only the case of  $G$ . The difference  $G_n(y; A, B) - G_n(y; A, C)$  equals

$$\frac{1}{1 - \zeta_n^-(C, A)} (F_n(y|A) - F_n(y|C)) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)).$$

Proceeding as in the proof to Theorem 6 immediately yields the asymptotic-distribution result that implies the theorem.  $\square$

## 4. SIMULATION EXPERIMENTS

In our numerical illustrations we will work with the family of skew-normal distributions (Azzalini 1985.) The skew-normal distribution with location  $\mu$ , positive scale  $\sigma$ , and



skewness parameter  $\beta$  multiplies the density of  $\mathcal{N}(\mu, \sigma^2)$  by a term that skews it to the right if  $\beta > 0$  and to the left if  $\beta < 0$ . Moreover,

$$f(x; \mu, \sigma, \beta) \equiv \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \times \frac{\Phi\left(\beta \frac{x - \mu}{\sigma}\right)}{\Phi(0)}.$$

Its mean and variance are  $\mu + \sigma\delta\sqrt{\frac{2}{\pi}}$  and  $\sigma^2\left(1 - \frac{2\delta^2}{\pi}\right)$ , respectively, where  $\delta \equiv \beta/\sqrt{1 + \beta^2}$ . Clearly,

$$f(x; \mu, \sigma, \beta) \rightarrow \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)$$

as  $\beta \rightarrow 0$ .

In our simulations we will consider data generating processes where the outcome is generated as

$$Y = T V_G + (1 - T) V_H, \quad (4.1)$$

where  $T$  is a latent binary variable, and  $V_G \sim G$  and  $V_H \sim H$ . Both error distributions  $G$  and  $H$  are skewed-normal distributions with parameters  $\mu_G, \sigma_G, \beta_G$  and  $\mu_H, \sigma_H, \beta_H$ , respectively.

From [Capitanio \(2010\)](#) it follows that Assumption 8 holds if  $G$  is right-skewed and  $H$  is left-skewed. We will consider designs where  $\beta_G > 0$  and  $\beta_H < 0$  to verify our asymptotics.

When  $\beta_G = \beta_H = 0$ , (4.1) collapses to a standard location model with normal errors

$$Y = (\mu_G - \mu_H) T + V, \quad V \sim \mathcal{N}(0, \sigma_G^2 + \sigma_H^2). \quad (4.2)$$

The identifying tail condition in Assumption 3 still holds if  $\mu_G > \mu_H$ , and our estimators remain consistent. However, Assumption 8 now fails and so we may expect poor inference in this design.

In our experiments we generate a binary  $X$  with  $\mathbb{P}(X = 1) = \frac{1}{2}$  and fix conditional probabilities as

$$\begin{aligned} \mathbb{P}(T = 0|X = 0) &= \frac{3}{4}, & \mathbb{P}(T = 1|X = 0) &= \frac{1}{4}, \\ \mathbb{P}(T = 1|X = 1) &= \frac{1}{4}, & \mathbb{P}(T = 0|X = 1) &= \frac{3}{4}. \end{aligned}$$

We present results for data generating processes where  $\mu_G = \mu = -\mu_H$  and  $\beta_G = \beta = -\beta_H$ . We use the designs  $\mu = 0$  and  $\beta \in \{2.5, 5\}$  to evaluate the adequacy of our

asymptotic arguments for small-sample inference. We also look at the performance of our estimators when  $\mu \in \{.5, 1\}$  and  $\beta = 0$ , which yields the Gaussian location model in (4.2). We fix  $\sigma_G = \sigma_H = 1$  throughout. For each of these designs we consider choices of the empirical quantiles as

$$\iota_{n_x} = C(n_x \ln \ln n_x)^{6/10}, \quad \kappa_{n_x} = C(n_x \ln \ln n_x)^{6/10},$$

for several choices of the constant  $C$ . All of these choices are in line with our asymptotic arguments. The larger the constant  $C$  the more conservative the choice of intermediate quantile,

$$q_\ell \equiv \frac{\iota_{n_x}}{n_x}, \quad q_r \equiv \frac{n_x - \kappa_{n_x}}{n_x},$$

for a given sample size.

We run experiments for sample sizes  $n \in \{500; 1,000, 2,500; 5,000; 10,000; 25,000\}$ . We report (the average over the replications of)  $q_\ell$  and  $q_r$  along with the estimation results to get an idea of how far in the tails of the component distributions we are going to obtain the results. A data-driven determination of the constant  $C$  is challenging and is left for future research. All tables and figures concerning our Monte Carlo experiment are collected at the end of the paper.

Tables 1–4 report the simulation results for the mixing proportions  $\lambda(0)$  and  $\lambda(1)$  for the data generating processes with  $\mu = 0, \beta = 5$ ;  $\mu = 0, \beta = 2.5$ ;  $\mu = 1, \beta = 0$ ; and  $\mu = .5, \beta = 0$ , respectively. Each table contains the bias, standard deviation (SD), ratio of the (average over the replications of the) estimated standard error to the standard deviation (SE/SD), and the coverage of 95% confidence intervals (CI95). All these statistics were computed from 10,000 Monte Carlo replications.

The results in Table 1 and Table 2 support our asymptotic theory. For each design and all choices of the tuning parameter  $C$ , the bias and standard deviation shrink to zero as  $n \uparrow +\infty$ ; and the bias is small relative to the standard error. Furthermore,  $\text{SE}/\text{SD} \rightarrow 1$  and the coverage rates of the confidence intervals are close to .95 in large samples. The variability of the point estimates is somewhat overestimated when  $n$  is very small and  $C$  is chosen conservatively. Together with the relatively small bias, this implies that confidence intervals are slightly conservative. For  $C = .5$ , coverage rates are close to .95,

even for the smallest samples considered, and for all  $C$ , the coverage rates move fairly quickly toward .95 as  $n$  increases.

Tables 3 and 4 present the results for the pure location model with Gaussian errors, where the tail conditions of Assumption 8 fail. The difference between the two designs is the distance between the component distributions. For Table 3,  $G$  is centered at 1 while  $H$  is centered at  $-1$ , so that  $\mu_G - \mu_H = 2$ . For Table 4,  $G$  and  $H$  are closer to each other:  $\mu_G - \mu_H = 1$ . In the first of these designs the bias in the point estimates is somewhat larger than in the skewed designs. Nonetheless, the bias is still small relative to the standard deviation. Furthermore, the coverage of the confidence intervals displays a similar pattern as in Tables 1 and 2, and is excellent for all values of  $C$  when  $n$  is not too small. When we move to Table 4 the bias increases further. The bias still shrinks to zero as  $n$  grows, confirming that our estimator remains consistent. However, the bias is not negligible relative to the standard deviation; the coverage of the confidence intervals deteriorates as  $n$  grows, and inference becomes unreliable.

We next turn to the results for the component distributions. For clarity we present the results by means of a series of plots. Figures 1–4 contain the results corresponding to Tables 1–4, respectively. To save on space we only present results for one value of  $C$ —the effect of larger values of  $C$  on  $G_n$  and  $H_n$  is similar to that on  $\lambda_n$  documented before. Each figure presents results for  $G_n$  (left plots) and  $H_n$  (right plots) for (from top to bottom)  $n = 1,000$ ,  $n = 2,500$ ,  $n = 5,000$ , and  $n = 10,000$ . Each plot contains the mean of the point estimates and the mean of 95% confidence bounds constructed using a plug-in estimator of the asymptotic variance in Theorem 5 (dashed lines). Each plot also contains the true component distribution and 95% confidence bounds constructed using the empirical standard deviation over the Monte Carlo replications (solid grey lines). We vary the range of the vertical axis across the plots in a given figure to enhance visibility.

Figures 1 and 2 again confirm our asymptotics. Indeed, moving down the rows,  $G_n$  and  $H_n$  converge to  $G$  and  $H$  respectively, while the estimated standard error converges to the Monte Carlo standard deviation. The bias in the point estimates is small across all plots. For small  $n$  the estimated standard errors are somewhat too small; this underestimation is more severe for  $H_n$  than for  $G_n$ .

Figures 3 and 4 concern the Gaussian location model. The results are in line with

our findings concerning the mixing proportions. In the design where  $\mu_G - \mu_H = 2$  our estimators do well in spite of Assumption 8 not holding. In Figure 4, however, the asymptotic bias in  $G_n$  and  $H_n$  becomes visible. While the variability of the point estimates is correctly captured by our asymptotic-variance estimator, the confidence bounds settle around an incorrect curve.

### CONCLUDING REMARKS

We conducted most of our analysis with a mixture of two components. However, some of our results would extend to a version of (1.1) with a larger number of components. Suppose that the mixture has  $J$  irreducible components, as in

$$F(y|x) = \sum_{j=1}^J \lambda_j(x) G_j(y),$$

in obvious notation. Henry et al. (2014) showed that the mixture components and mixing proportions are only identified up to  $J(J-1)$  inequality-constrained real parameters in general.

Tail dominance restrictions can still be quite powerful. Take  $J = 3$  for instance, and assume that  $G_1$  dominates in the left tail and  $G_3$  dominates in the right tail. Then it is easy to adapt the proof of Theorem 1 to prove that the behavior of  $F(y|x)$  in the left tail identifies the function  $\lambda_1$  up to a multiplicative constant, and that the behavior of  $F(y|x)$  in the right tail identifies the function  $\lambda_3$  up to another multiplicative constant. Imposing the values of the mixing proportions at one particular value of  $x$  would be enough to point identify all elements of the model, for instance; and it would be easy to adapt our estimators and tests to such a setting. Whether such additional restrictions are plausible is, of course, highly model-dependent.

**Table 1.**  $\mu = 0$  and  $\beta = 5$ 

$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0115	−.0083	.0951	.0914	1.0790	1.0384	.9714	.9694
1,000	.059	.940	.0060	−.0059	.0693	.0701	1.0554	1.0392	.9688	.9682
2,500	.043	.957	.0038	−.0029	.0510	.0506	1.0321	1.0393	.9624	.9636
5,000	.034	.966	.0016	−.0023	.0406	.0407	1.0246	1.0246	.9588	.9592
10,000	.026	.974	.0012	−.0011	.0328	.0325	1.0106	1.0213	.9560	.9572
25,000	.018	.982	.0013	−.0007	.0245	.0244	1.0108	1.0149	.9526	.9568
$C = 1$										
500	.172	.848	.0044	−.0046	.0534	.0535	1.1937	1.1907	.9904	.9876
1,000	.120	.880	.0024	−.0035	.0439	.0446	1.1358	1.1220	.9764	.9752
2,500	.087	.914	.0012	−.0010	.0332	.0337	1.1031	1.0847	.9680	.9668
5,000	.067	.933	.0012	−.0005	.0270	.0271	1.0814	1.0726	.9664	.9670
10,000	.052	.947	.0007	−.0003	.0225	.0222	1.0360	1.0519	.9566	.9616
25,000	.037	.963	.0006	−.0005	.0167	.0167	1.0459	1.0460	.9594	.9600
$C = 1.5$										
500	.227	.773	.0198	−.0198	.0328	.0333	1.6715	1.6482	.9982	.9984
1,000	.179	.821	.0046	−.0037	.0316	.0315	1.2931	1.2933	.9944	.9920
2,500	.129	.870	.0013	.0000	.0257	.0256	1.1575	1.1579	.9774	.9762
5,000	.101	.898	.0010	−.0005	.0208	.0215	1.1427	1.1070	.9756	.9720
10,000	.078	.922	.0002	−.0010	.0175	.0174	1.0873	1.0962	.9646	.9710
25,000	.056	.944	.0004	.0000	.0136	.0136	1.0453	1.0474	.9632	.9604

**Table 2.**  $\mu = 0$  and  $\beta = 2.5$ 

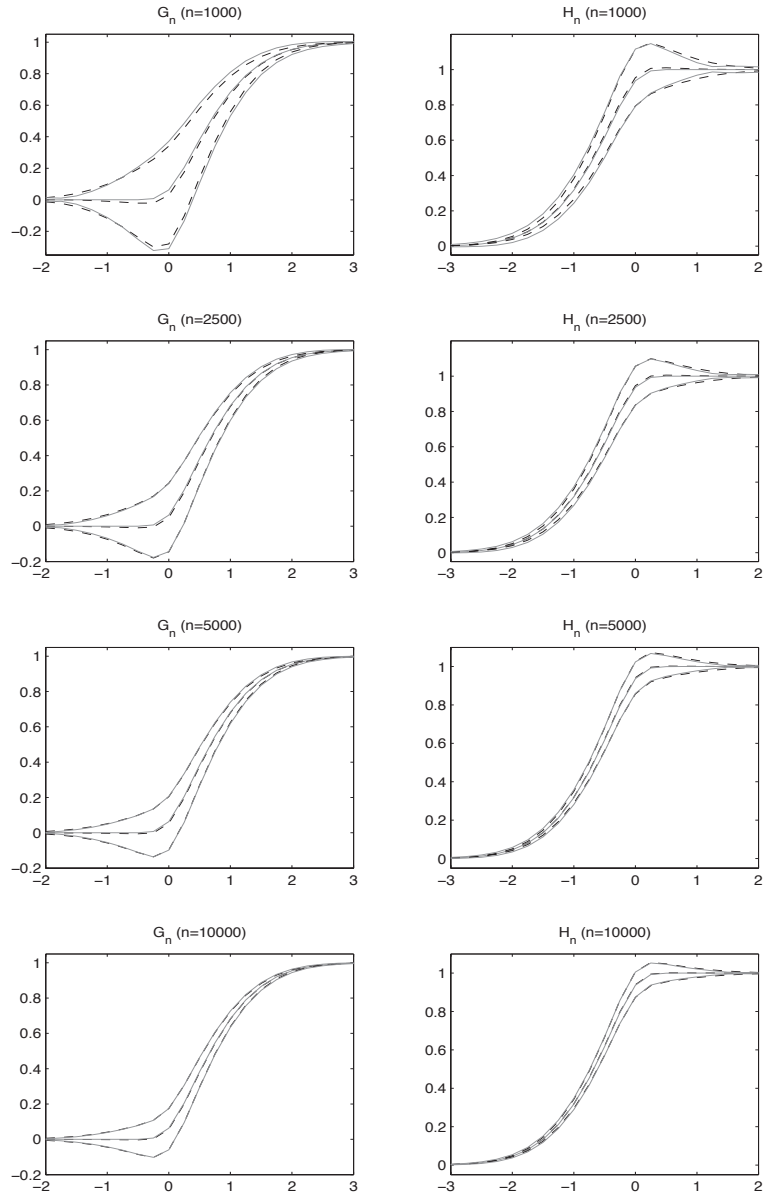
	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0093	−.0122	.0941	.0934	1.0161	1.0315	.9718	.9780
1,000	.059	.940	.0066	−.0072	.0722	.0718	1.0151	1.0194	.9646	.9652
2,500	.043	.957	.0029	−.0040	.0512	.0519	1.0279	1.0212	.9610	.9612
5,000	.034	.966	.0016	−.0026	.0407	.0407	1.0248	1.0255	.9546	.9596
10,000	.026	.974	.0012	−.0015	.0323	.0326	1.0287	1.0193	.9548	.9626
25,000	.018	.982	.0007	−.0009	.0243	.0244	1.0182	1.0137	.9564	.9580
$C = 1$										
500	.172	.848	.0160	−.0174	.0519	.0518	1.2924	1.2969	.9954	.9972
1,000	.120	.880	.0069	−.0084	.0429	.0430	1.1865	1.1898	.9862	.9890
2,500	.087	.914	.0032	−.0024	.0334	.0338	1.1035	1.0900	.9716	.9684
5,000	.067	.933	.0018	−.0014	.0270	.0275	1.0868	1.0648	.9700	.9666
10,000	.052	.947	.0001	−.0010	.0220	.0219	1.0569	1.0661	.9636	.9652
25,000	.037	.963	.0005	−.0005	.0168	.0168	1.0376	1.0356	.9603	.9585
$C = 1.5$										
500	.227	.773	.0491	−.0496	.0361	.0365	1.7108	1.6925	.9908	.9902
1,000	.179	.821	.0245	−.0248	.0304	.0303	1.4562	1.4670	.9920	.9926
2,500	.129	.870	.0083	−.0080	.0248	.0250	1.2350	1.2249	.9868	.9872
5,000	.101	.898	.0030	−.0035	.0211	.0210	1.1386	1.1418	.9774	.9788
10,000	.078	.922	.0015	−.0016	.0172	.0172	1.1103	1.1090	.9722	.9720
25,000	.056	.944	.0005	−.0007	.0136	.0136	1.0496	1.0455	.9604	.9568

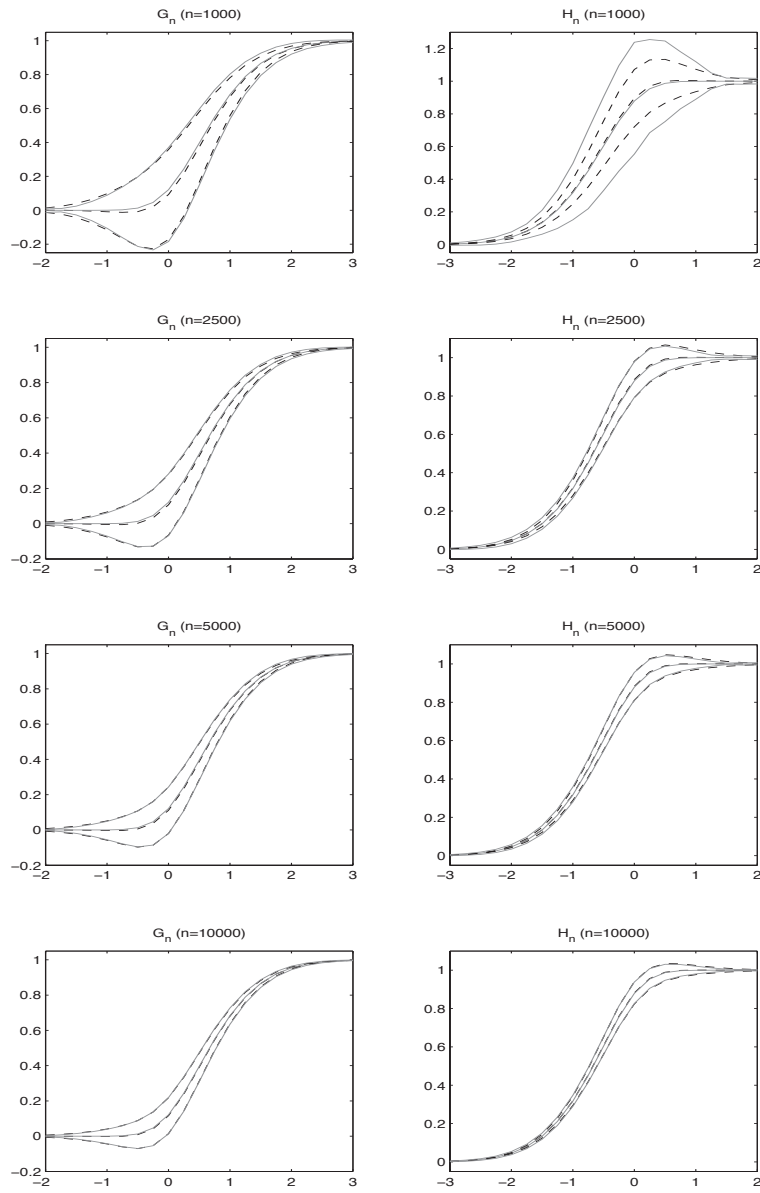
**Table 3.**  $\mu = 1$  and  $\beta = 0$ 

$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0220	−.0227	.0966	.0948	1.0479	1.0710	.9869	.9866
1,000	.059	.940	.0144	−.0164	.0720	.0728	1.0589	1.0518	.9807	.9810
2,500	.043	.957	.0092	−.0105	.0519	.0527	1.0476	1.0331	.9705	.9697
5,000	.034	.966	.0067	−.0069	.0411	.0415	1.0372	1.0283	.9673	.9651
10,000	.026	.974	.0050	−.0048	.0327	.0324	1.0344	1.0449	.9614	.9622
25,000	.018	.982	.0030	−.0033	.0246	.0245	1.0188	1.0246	.9576	.9586
$C = 1$										
500	.172	.848	.0385	−.0395	.0576	.0575	1.2942	1.2958	.9948	.9944
1,000	.120	.880	.0268	−.0267	.0459	.0453	1.2162	1.2304	.9867	.9869
2,500	.087	.914	.0154	−.0166	.0339	.0344	1.1554	1.1417	.9786	.9770
5,000	.067	.933	.0110	−.0120	.0274	.0278	1.1168	1.1036	.9722	.9691
10,000	.052	.947	.0079	−.0084	.0225	.0224	1.0752	1.0823	.9641	.9656
25,000	.037	.963	.0055	−.0055	.0169	.0173	1.0602	1.0361	.9603	.9562
$C = 1.5$										
500	.227	.773	.0718	−.0714	.0421	.0418	1.6251	1.6382	.9726	.9739
1,000	.179	.821	.0476	−.0474	.0343	.0340	1.4396	1.4479	.9619	.9632
2,500	.129	.870	.0277	−.0282	.0269	.0270	1.2510	1.2445	.9502	.9487
5,000	.101	.898	.0188	−.0191	.0217	.0219	1.1929	1.1831	.9510	.9473
10,000	.078	.922	.0133	−.0132	.0178	.0179	1.1330	1.1285	.9443	.9469
25,000	.056	.944	.0086	−.0084	.0135	.0136	1.0949	1.0895	.9424	.9458

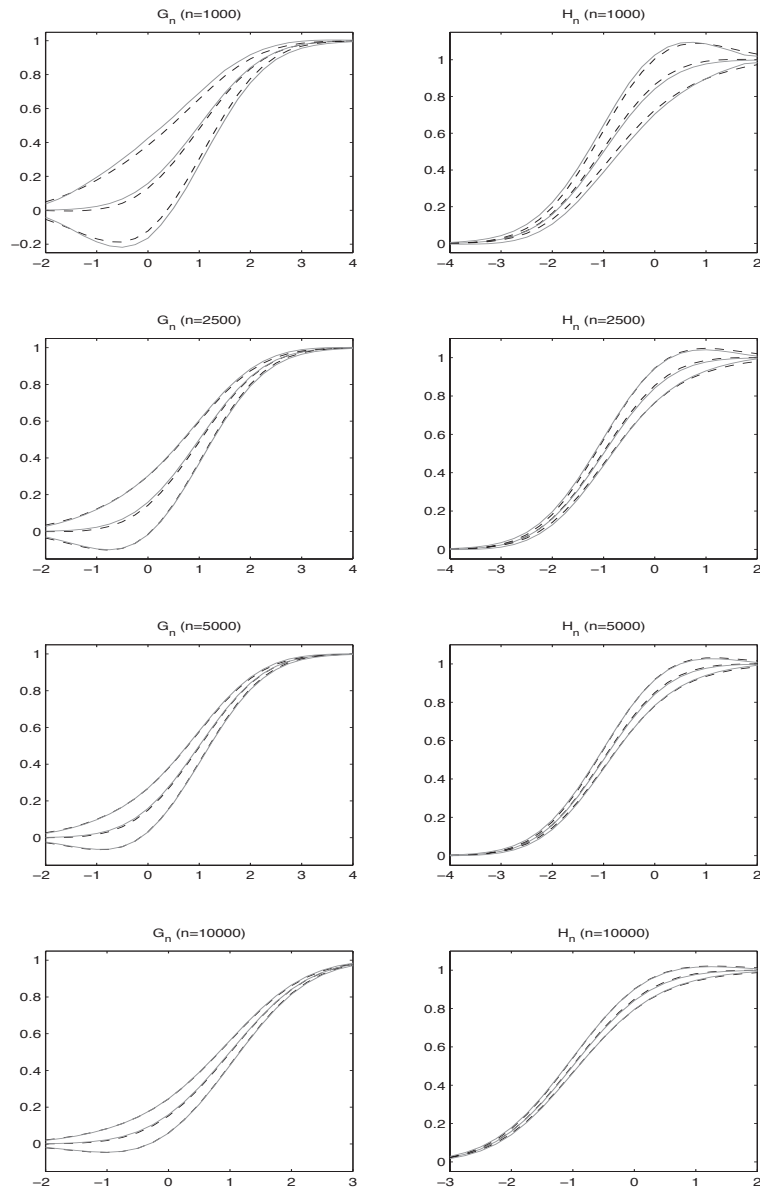
**Table 4.**  $\mu = .5$  and  $\beta = 0$ 

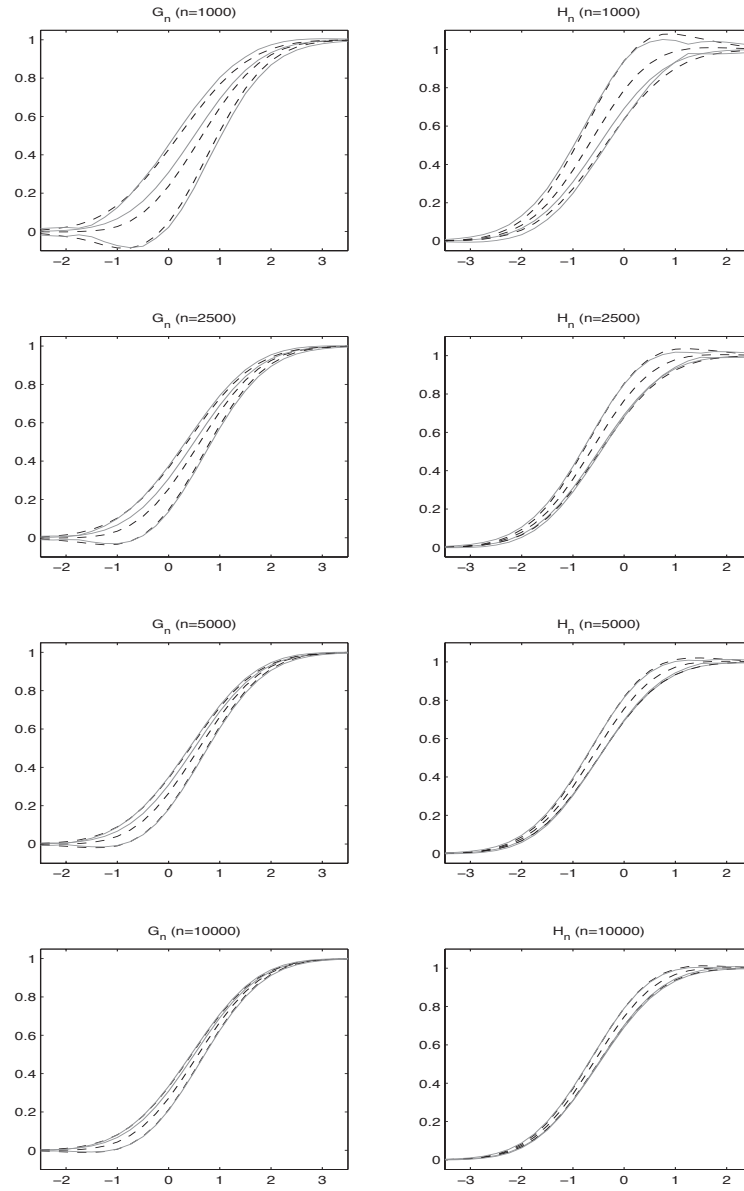
	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .75$										
500	.114	.886	.1131	−.1145	.1162	.1147	1.1992	1.2238	.9864	.9880
1,000	.090	.910	.0994	−.1017	.0842	.0855	1.1677	1.1599	.9416	.9406
2,500	.065	.935	.0840	−.0864	.0594	.0607	1.1175	1.0999	.8470	.8404
5,000	.051	.949	.0755	−.0758	.0454	.0463	1.1092	1.0920	.7448	.7342
10,000	.039	.961	.0671	−.0671	.0358	.0352	1.0815	1.0973	.6244	.6286
25,000	.028	.972	.0580	−.0573	.0257	.0259	1.0710	1.0601	.4374	.4584
$C = 1.25$										
500	.189	.811	.1451	−.1445	.0905	.0915	1.3579	1.3417	.9196	.9150
1,000	.149	.851	.1258	−.1271	.0673	.0682	1.2780	1.2632	.7896	.7808
2,500	.108	.892	.1075	−.1065	.0476	.0475	1.2051	1.2037	.5524	.5472
5,000	.084	.916	.0938	−.0947	.0369	.0371	1.1590	1.1555	.3806	.3672
10,000	.065	.935	.0832	−.0837	.0284	.0284	1.1426	1.1407	.2146	.2208
25,000	.047	.953	.0713	−.0714	.0205	.0208	1.1143	1.1002	.0740	.0740
$C = 1.75$										
500	.265	.735	.1704	−.1723	.0747	.0757	1.5656	1.5429	.8094	.7980
1,000	.209	.791	.1498	−.1500	.0573	.0580	1.4223	1.4072	.5592	.5578
2,500	.152	.849	.1255	−.1253	.0408	.0417	1.2970	1.2674	.2658	.2702
5,000	.118	.882	.1114	−.1095	.0323	.0318	1.2143	1.2267	.1076	.1160
10,000	.092	.908	.0968	−.0980	.0247	.0248	1.1889	1.1849	.0358	.0296
25,000	.065	.935	.0827	−.0827	.0178	.0180	1.1481	1.1324	.0034	.0030

**Figure 1.**  $\mu = 0$  and  $\beta = 5$  ( $C = .5$ )

**Figure 2.**  $\mu = 0$  and  $\beta = 2.5$  ( $C = .5$ )



**Figure 3.**  $\mu = 1$  and  $\beta = 0$  ( $C = .5$ )

**Figure 4.**  $\mu = .5$  and  $\beta = 0$  ( $C = .75$ )

## REFERENCES

- Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tabaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica* 80(5), 1977–2016.
- Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* 37, 3099–3132.
- Andrews, D. W. K. and M. M. A. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.
- Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review* 102, 94–130.
- Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature* 49, 3–71.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.
- Bollinger, C. R. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics* 73, 387–399.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014a). Estimating multivariate latent-structure models. Sciences Po Discussion Paper 2014-18.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014b). Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society - Series B*, forthcoming.
- Bordes, L., S. Mottelet, and P. Vandekerckhove (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics* 34, 1204–1232.
- Capitanio, A. (2010). On the approximation of the tail probability of the scalar skew-normal distribution. *METRON* 68, 299–308.
- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall, CRC Press.
- D’Haultfoeuille, X. and P. Février (2014). Identification of mixture models using support variations. Mimeo.
- D’Haultfoeuille, X. and A. Maurel (2013). Another look at identification at infinity of sample selection models. *Econometric Theory* 29, 213–224.
- Einmahl, J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *Journal of Statistical Planning and Inference* 32, 137–145.
- Einmahl, U. and D. Mason (1997). Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields* 107, 283–311.
- Frisch, R. (1934). Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics* 1, 255–294.
- Gassiat, E. and J. Rousseau (2014). Non parametric finite translation mixtures with dependent regime. *Bernoulli*, forthcoming.
- Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics Volume 14: Statistical Methods in Finance*. Elsevier.

- Hall, P. and X.-H. Zhou (2003). Nonparametric identification of component distributions in a multivariate mixture. *Annals of Statistics* 31, 201–224.
- Hamilton, J. D. (1989). A new approach to the analysis of nonstationary times series and the business cycle. *Econometrica* 57, 357–384.
- Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.
- Heckman, J. J. (1990). Varieties of selection bias. *American Economic Review* 80, 313–318.
- Henry, M., Y. Kitamura, and B. Salanié (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Paper 1767.
- Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5, 123–144.
- Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.
- Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77, 135–175.
- Khan, S. and E. Tamer (2010). Irregular identification, support conditions and inverse weight estimation. *Econometrica* 78, 2021–2042.
- Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.
- Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica* 74, 631–665.
- Schwarz, M. and S. Van Belleghem (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters* 80, 236–241.
- Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica* 68, 343–369.

## INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS\*

KOEN JOCHMANS<sup>1</sup> and MARC HENRY<sup>2</sup> and BERNARD SALANIÉ<sup>3</sup>

<sup>1</sup>*Sciences Po, 28 rue des Saint Pères, 75007 Paris, France.*

*E-mail:* [koen.jochmans@sciencespo.fr](mailto:koen.jochmans@sciencespo.fr)

<sup>2</sup>*The Pennsylvania State University, University Park, PA 16801, U.S.A.*

*E-mail:* [marc.henry@psu.edu](mailto:marc.henry@psu.edu)

<sup>3</sup>*Columbia University, 420 West 118th Street, New York, NY 10027, U.S.A.*

*E-mail:* [bsalanie@columbia.edu](mailto:bsalanie@columbia.edu)

Revised on March 27, 2015

Many econometric models can be analyzed as finite mixtures. We focus on two-component mixtures and we show that they are nonparametrically point identified by a combination of an exclusion restriction and tail restrictions. Our identification analysis suggests simple closed-form estimators of the component distributions and mixing proportions. We derive their asymptotic properties using results on tail empirical processes and we present a simulation study that documents their finite-sample performance.

*Keywords:* mixture model, nonparametric identification and estimation, tail empirical process.

## INTRODUCTION

The use of finite mixtures has a long history in applied econometrics. A non-exhaustive list of applications includes models with discrete unobserved heterogeneity, hidden Markov chains, and models with mismeasured discrete variables; see [Henry et al. \(2014\)](#) for a more extensive discussion of applications. Until recently, the literature on nonparametric identification of mixture models was sparse. Following the lead of [Hall and Zhou \(2003\)](#), several authors have analyzed multivariate mixtures; recent contributions are [Kasahara and Shimotsu \(2009\)](#), [Allman et al. \(2009\)](#), and [Bonhomme et al. \(2014a,b\)](#). There are fewer identifying restrictions available when the model of interest is univariate. [Bordes](#)

---

\*We are grateful to Arthur Lewbel and three referees for comments and suggestions, and to Victor Chernozhukov and Yuichi Kitamura for fruitful discussions. Parts of this paper were written while Henry was visiting the University of Tokyo Graduate School of Economics and while Salanié was visiting the Toulouse School of Economics. The hospitality of both institutions is gratefully acknowledged. Henry's research has received funding from SSHRC Grants 410-2010-242 and 435-2013-0292, and NSERC Grant 356491-2013. Jochmans' research has received funding from the SAB grant "Nonparametric estimation of finite mixtures". Some of the results presented here previously circulated as part of [Henry et al. \(2010\)](#), whose published version ([Henry et al. 2014](#)) only contains results on partial identification.

et al. (2006), for instance, provide such restrictions for location models with symmetric error distributions.

In this paper we give sufficient conditions that point-identify univariate component distributions and associated mixing proportions. The restrictions we rely on are most effective in two-component models. To simplify the analysis we focus on this case, like Hall and Zhou (2003) and Bordes et al. (2006). We comment briefly on mixtures with more components at the end of the paper. Our arguments are constructive, and we propose closed-form estimators for both the component distributions and the mixing proportions. We derive their large-sample properties and investigate their behavior in a simulation experiment.

The model we consider in this paper is characterized by an exclusion restriction and a tail-dominance assumption. Like Henry et al. (2014), we assume the existence of a source of variation that shifts the mixing proportions but leaves the component distributions unchanged. Such an assumption is natural in several important applications, such as measurement-error models (Mahajan 2006). In hidden Markov models, it follows directly from the model specification. The exclusion restriction is also implied by the conditional-independence restriction that underlies the results of Hall and Zhou (2003) and others on multivariate mixtures.

It was shown in Henry et al. (2014) that our exclusion restriction restricts both the component distributions and the mixing proportions to lie in a non-trivial set. Here we achieve point-identification by complementing the exclusion restriction with a restriction on the relative tail behavior of the component distributions. This restriction is quite natural in location models, for instance, but it can be motivated more generally. For example, regime-switching models typically feature regimes with different tail behavior. Alternatively, theoretical models can imply the required tail behavior; an example is the search and matching model of Shimer and Smith (2000), as explained in D'Haultfoeuille and Février (2014).

Our identification argument suggests plug-in estimators of the mixing proportions and the component distributions that are available in closed form. The estimators are based on ratios of intermediate quantiles, and their convergence rate is determined by the theory of tail empirical processes. As we rely on the tail behavior of the component

distributions to infer the mixing proportions, our estimators converge more slowly than the parametric rate. If the mixing proportions were known—or could be estimated at the parametric rate—the tail restrictions could be dispensed with and the implied estimator of the component distributions would also converge at the parametric rate.

Our estimators are consistent under very weak tail dominance assumptions. To control for asymptotic bias in their limit distribution we need to impose stronger requirements that prevent the tails of the components from vanishing too quickly. These assumptions rule out the Gaussian location model. Such thin-tailed distributions are known to be problematic for inference techniques that rely on tail behavior ([Khan and Tamer 2010](#)). However, we show that our assumptions apply to distributions with fatter tails, such as Pareto distributions.

Identification only requires that the variable subject to the exclusion restriction can take on two values. If it can take on more values, the model is overidentified and the specification can be tested. We derive the asymptotic properties of two statistics that can be used to respectively test the validity of the tail restrictions and that of the exclusion restriction.

The tail conditions we use to obtain nonparametric identification are related to the well-known identification-at-infinity argument of [Heckman \(1990\)](#); see also [D’Haultfoeulle and Maurel \(2013\)](#) for another approach. Other types of support restrictions have been used in related problems to establish identification. [Schwarz and Van Bellegem \(2010\)](#) imposed support restrictions in a semiparametric deconvolution problem to deal with measurement error in location models. [D’Haultfoeulle and Février \(2014\)](#) relied on a support condition as an alternative to completeness conditions ([Hu and Schennach 2008](#)) in multivariate mixture models.

The remainder of the paper is organized as follows. Section 1 describes the mixture model and proves identification. We rely on these results to construct estimators and derive their asymptotic properties in Section 2. Section 3 describes our specification tests. In Section 4 we conduct a Monte Carlo experiment that gives evidence on the small-sample performance of our methods. Finally, we conclude with some remarks on mixtures with more than two components.

## 1. MIXTURES WITH EXCLUSION AND TAIL RESTRICTIONS

Let  $(Y, X) \in \mathbb{R} \times \mathcal{X}$  be random variables. We assume throughout that our mixtures satisfy the following simple exclusion restriction.<sup>1</sup>

ASSUMPTION 1 (Mixture with exclusion).  *$F(y|x) \equiv \mathbb{P}(Y \leq y|X = x)$  decomposes as the two-component mixture*

$$F(y|x) = G(y) \lambda(x) + H(y) (1 - \lambda(x)) \quad (1.1)$$

*for distribution functions  $G : \mathbb{R} \mapsto [0, 1]$  and  $H : \mathbb{R} \mapsto [0, 1]$  and a function  $\lambda : \mathcal{X} \mapsto [0, 1]$  that maps values  $x$  into mixing proportions.*

The assumption that the component distributions do not depend on  $X$  embodies our exclusion restriction; see also Henry et al. (2014).

We complete the mixture model with the following assumption.

ASSUMPTION 2. *The mixing proportion  $\lambda$  is non-constant on  $\mathcal{X}$  and is bounded away from zero and one on  $\mathcal{X}$ .*

Non-constancy of  $\lambda$  gives the variable  $X$  relevance. Bounding  $\lambda$  away from zero and one implies that the mixture is irreducible.<sup>2</sup>

### 1.1. Motivating examples

Our first example has a long history in empirical work (Frisch 1934).

EXAMPLE 1 (Mismeasured treatments). *Let  $T$  denote a binary treatment indicator. Suppose that  $T$  is subject to classification error: rather than observing  $T$ , we observe misclassified treatment  $X$ . The distribution of the outcome variable  $Y$  given  $X = x$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1, X = x) \lambda(x) + \mathbb{P}(Y \leq y|T = 0, X = x) (1 - \lambda(x)),$$

---

<sup>1</sup>We omit conditioning variables throughout. The identification analysis extends straightforwardly. In principle, the distribution theory could be extended by using local empirical process results along the lines of Einmahl and Mason (1997). We postpone a detailed investigation into such an extension to future work.

<sup>2</sup>Note that irreducibility rules out the possibility of achieving identification of  $G$  and  $H$  via an identification-at-infinity argument, as in Heckman (1990) and Andrews and Schafgans (1998) for instance.



with  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . The usual ignorability assumption states that  $X$  and  $Y$  are independent given  $T$ . That is,

$$\mathbb{P}(Y \leq y|T = t, X = x) = \mathbb{P}(Y \leq y|T = t),$$

for  $t \in \{0, 1\}$ , in which case the decomposition of  $F(y|x)$  reduces to the model in (1.1) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$  and  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ . Also note that  $\lambda$  is non-constant unless misclassification in  $T$  is completely random.

The identification of treatment effects when the treatment indicator is mismeasured has received considerable attention, especially in the context of regression models (Bollinger 1996; Mahajan 2006; Lewbel 2007). Here, the conditional ignorability assumption that validates our exclusion restriction relies on non-differential misclassification error. It has been routinely used elsewhere (Carroll et al. 2006).

Our second example deals with regime-switching models, also referred to as hidden Markov models. These models cover switching regressions, which have been used in a variety of settings (see, e.g., Heckman 1974, Hamilton 1989), as well as several versions of stochastic-volatility models (Ghysels et al. 1996).

EXAMPLE 2 (Hidden Markov model). *Let  $\mathbf{Y} = (Y_1, \dots, Y_T)'$  be a time series of outcome variables. A hidden Markov model for the dependency structure in these data assumes that there is a discrete latent series of state variables  $\mathbf{S} = (S_1, \dots, S_T)'$  having Markov dependence, that the variables in  $\mathbf{Y}$  are jointly independent given  $\mathbf{S}$ , and that*

$$\mathbb{P}(Y_t \leq y_t|\mathbf{S} = \mathbf{s}) = \mathbb{P}(Y_t \leq y_t|S_t = s_t).$$

*To see that such a model fits (1.1), assume that there are two latent states 0 and 1 and (for notational simplicity) that  $\mathbf{S}$  has first-order Markov dependence. Denote  $\mathbf{X} = (Y_1, \dots, Y_{t-1})'$ . Then*

$$F(y_t|\mathbf{x}) = \mathbb{P}(Y_t \leq y_t|S_t = 1)\mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x}) + \mathbb{P}(Y_t \leq y_t|S_t = 0)\mathbb{P}(S_t = 0|\mathbf{X} = \mathbf{x}),$$

*which fits our setup. Moreover,  $\lambda(\mathbf{x}) = \mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x})$  does vary with  $\mathbf{x}$ , unless the outcomes are independent of the latent states.*

In this example, the exclusion restriction follows directly from the Markovian structure of the regime-switching model. [Gassiat and Rousseau \(2014\)](#) obtained nonparametric identification in location models when the matrix of transition probabilities of the Markov chain has full rank. The approach presented here delivers nonparametric identification in a much broader range of models.

Our third example links [\(1.1\)](#) to the recent literature on multivariate mixtures building on [Hall and Zhou \(2003\)](#).

**EXAMPLE 3** (Multivariate mixture). *Suppose  $Y$  and  $X$  are two measurements that are independent conditional on a latent binary factor  $T$ :*

$$\begin{aligned}\mathbb{P}(Y \leq y, X \leq x) &= \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(X \leq x|T = 1) \mathbb{P}(T = 1) \\ &\quad + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(X \leq x|T = 0) \mathbb{P}(T = 0).\end{aligned}$$

*Then the conditional distribution of the  $Y$  given  $X$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(T = 1|X = x) + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(T = 0|X = x).$$

*This is of the form in [\(1.1\)](#) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$ ,  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ , and  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . Note that the bivariate mixture model implies that the distribution of  $X$  given  $Y$  decomposes in the same way.*

[Hall and Zhou \(2003\)](#) showed that multivariate two-component mixtures with conditional independence restrictions are nonparametrically identified from data on three or more measurements and are set identified from data on only two measurements. The results we derive below imply that two measurements can also yield point identification under tail restrictions.

## 1.2. Identification

We show below that both the mixture components  $G, H$  and the mixing proportions  $\lambda$  are identified under the following dominance condition on the tails of the component distributions.

ASSUMPTION 3 (Tail dominance).

(i) The left tail of  $G$  is thinner than the left tail of  $H$ , i.e.,

$$\lim_{y \downarrow -\infty} \frac{G(y)}{H(y)} = 0.$$

(ii) The right tail of  $G$  is thicker than the right tail of  $H$ , i.e.,

$$\lim_{y \uparrow +\infty} \frac{1 - H(y)}{1 - G(y)} = 0.$$

Tail dominance is natural in location models.

EXAMPLE 4 (Location models). Suppose that  $Y = \mu(T) + U$ , where  $T$  is a binary indicator and  $U \sim F$ , independent of  $T$ . Then (1.1) yields

$$F(y|x) = F(y - \mu(1)) \mathbb{P}(T = 1|X = x) + F(y - \mu(0)) \mathbb{P}(T = 0|X = x).$$

Suppose that  $\mu(0) < \mu(1)$ , that  $F$  is absolutely continuous, and that the derivatives of  $-\ln(1 - F(u))$  and  $-\ln F(u)$  grow without bound as  $u \uparrow +\infty$  and  $u \downarrow -\infty$ , respectively. Then Assumption 3 holds with  $G(y) = F(y - \mu(1))$  and  $H(y) = F(y - \mu(0))$ .

**Proof.** We prove Assumption 3(ii). Let  $\varphi(u) \equiv -\ln(1 - F(u))$  and denote  $\varphi'$  its first derivative. Then

$$\frac{1 - F(y - \mu(0))}{1 - F(y - \mu(1))} = \exp(\varphi(y - \mu(1)) - \varphi(y - \mu(0))) = \exp(-\varphi'(y^*)(\mu(1) - \mu(0))),$$

where  $y^*$  lies between  $y - \mu(1)$  and  $y - \mu(0)$ . Since  $\varphi'$  increases without bound as  $y \uparrow +\infty$  and  $\mu(1) > \mu(0)$ , the expression on the right-hand side tends to zero as  $y$  increases. Assumption 3(i) can be verified in the same way.  $\square$

We now show that, combined, our exclusion restriction and tail dominance identify all elements of the mixture model.

THEOREM 1 (Identification). Under Assumptions 1–3,  $G$ ,  $H$ , and  $\lambda$  are identified.

**Proof.** The proof is constructive. Fix  $x' \in \mathcal{X}$  and choose  $x'' \in \mathcal{X}$  so that  $\lambda(x') \neq \lambda(x'')$ . Then re-arranging (1.1) gives

$$\begin{aligned} \frac{F(y|x')}{F(y|x'')} &= \frac{1 + \lambda(x') (G(y)/H(y) - 1)}{1 + \lambda(x'') (G(y)/H(y) - 1)}, \\ \frac{1 - F(y|x')}{1 - F(y|x'')} &= \frac{\lambda(x') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x'))}{\lambda(x'') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x''))}. \end{aligned}$$

Taking limits, Assumption 3 further implies that

$$\begin{aligned} \zeta^-(x', x'') &\equiv \lim_{y \downarrow -\infty} \frac{F(y|x')}{F(y|x'')} = \frac{1 - \lambda(x')}{1 - \lambda(x'')}, \\ \zeta^+(x', x'') &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|x')}{1 - F(y|x'')} = \frac{\lambda(x')}{\lambda(x'')}. \end{aligned} \tag{1.2}$$

These two equations can be solved for the mixing proportion at  $x'$ , yielding

$$\lambda(x') = \frac{1 - \zeta^-(x'', x')}{\zeta^+(x'', x') - \zeta^-(x'', x')}. \tag{1.3}$$

Since  $\lambda$  is non-constant, for any  $x' \in \mathcal{X}$  there exists a  $x'' \in \mathcal{X}$  for which such a system of equations can be constructed. The function  $\lambda$  is therefore identified on its entire support.

To establish identification of  $G$  and  $H$ , first note that

$$G(y) - H(y) = \frac{F(y|x'') - F(y|x')}{\lambda(x'') - \lambda(x')} \tag{1.4}$$

follows from (1.1). Then, evaluating (1.1) in  $x''$  and re-arranging the resulting expression for  $F(y|x'')$  gives

$$H(y) = F(y|x'') - (G(y) - H(y)) \lambda(x'') = F(y|x'') - \frac{\lambda(x'')}{\lambda(x'') - \lambda(x')} (F(y|x'') - F(y|x')),$$

which is identified. Furthermore, using (1.2) we can write

$$H(y) = F(y|x'') - \frac{1}{1 - \zeta^+(x', x'')} (F(y|x'') - F(y|x')). \tag{1.5}$$

Plugging this expression for  $H(y)$  back into the mixture representation of  $F(y|x'')$  as in (1.1) further yields

$$G(y) = F(y|x'') - \frac{1}{1 - \zeta^-(x', x'')} (F(y|x'') - F(y|x')), \tag{1.6}$$

again using (1.2). This shows that both component distributions are identified, concluding the proof.  $\square$

If we only assume one-sided tail dominance, then either  $G$  or  $H$  remains identified.

**COROLLARY 1** (One-sided tail dominance). *Under Assumptions 1 and 2,  $G$  is identified if Assumption 3(i) holds and  $H$  is identified if Assumption 3(ii) holds.*

**Proof.** We consider identification of  $H$ . Let  $x', x''$  be as in the proof of Theorem 1. Under Assumption 3(ii) we can still determine  $\zeta^+(x', x'') = \lambda(x')/\lambda(x'')$ , from which we can learn the ratio  $1/(1 - \zeta^+(x'', x'))$ . Together with (1.5) this yields  $H$ . This concludes the proof of the corollary.  $\square$

The following example illustrates the usefulness of Corollary 1.

**EXAMPLE 5** (Stochastic volatility). *Consider a two-regime stochastic volatility model, which is a special case of Example 2. Assume that the outcome variable  $Y$  has mean zero and conditional variance*

$$T\sigma_G^2 + (1 - T)\sigma_H^2$$

*for positive constants  $\sigma_G^2$  and  $\sigma_H^2$ . Suppose that  $\sigma_G^2 > \sigma_H^2$ . Then  $G$  is the distribution associated with a regime that is characterized by relatively higher volatility. In this case, both tails of  $G$  dominate those of  $H$ . Hence, in Assumption 3, Condition (ii) holds but Condition (i) fails. Nevertheless, the distribution of the lower-volatility regime, namely  $H$ , remains identified.*

Our identification result suggests plug-in estimators of the mixing proportions and the component distributions. The proof of Theorem 1, and Equations (1.3) and (1.5)–(1.6) in particular, further show that our mixture model yields overidentifying restrictions as soon as the instrument can take on more than two values. We turn to estimation in the next section. Afterwards we discuss the construction of specification tests under overidentification.

## 2. ESTIMATION

To motivate the construction of our estimators, we first note that the structure of the model in (1.1) continues to hold when we aggregate across  $x$ . Extending our notation to

$$F(y|A) \equiv \mathbb{P}(Y \leq y|X \in A), \quad \lambda(A) \equiv \sum_{x \in A} \lambda(x) \mathbb{P}(X = x|X \in A),$$

for any  $A \subset \mathcal{X}$ , we have

$$F(y|A) = G(y) \lambda(A) + H(y) (1 - \lambda(A)), \quad (2.1)$$

which is of the same form as (1.1). Furthermore, the proof of Theorem 1 continues to go through for (2.1); replacing  $x'$  with  $A$  and  $x''$  with  $\mathcal{X} - A$  does not alter the argument.

We will assume from now on that  $X$  is discrete. As will become apparent, this only entails a loss of generality for the estimation of the function  $\lambda$ , as our estimator will only yield a discretized approximation to it. Extending our results to continuous  $X$  would complicate the exposition greatly and we feel that it would only distract from our main argument.

We will work under the following sampling condition.

ASSUMPTION 4.  $(Y_1, X_1), \dots, (Y_n, X_n)$  is a random sample on  $(Y, X)$ .

For each  $A \subset \mathcal{X}$ , let

$$F_n(y|A) \equiv n_A^{-1} \sum_{i=1}^n 1\{Y_i \leq y, X_i \in A\},$$

where  $n_A \equiv \sum_{i=1}^n 1\{X_i \in A\}$ .

For each pair of disjoint subsets  $A, B$  of  $\mathcal{X}$  we can generalize (1.2) to

$$\begin{aligned} \zeta^-(A, B) &\equiv \lim_{y \downarrow -\infty} \frac{F(y|A)}{F(y|B)} = \frac{1 - \lambda(A)}{1 - \lambda(B)}, \\ \zeta^+(A, B) &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|A)}{1 - F(y|B)} = \frac{\lambda(A)}{\lambda(B)}. \end{aligned} \quad (2.2)$$

For any subsample of size  $m$  and integers  $\iota_m$  and  $\kappa_m$ , let  $\ell_m$  and  $r_m$  denote the  $(\iota_m + 1)$ th and  $(m - \kappa_m)$ th order statistics of  $Y$  in this subsample. We estimate the quantities in (2.2) by

$$\zeta_n^-(A, B) \equiv \frac{F_n(\ell_{n_B}|A)}{F_n(\ell_{n_B}|B)}, \quad \zeta_n^+(A, B) \equiv \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)}, \quad (2.3)$$

respectively. In our asymptotic theory, we will choose  $\iota_{n_B}$  and  $\kappa_{n_B}$  so that  $\ell_{n_B} \downarrow -\infty$  and  $r_{n_B} \uparrow +\infty$  as  $n \uparrow +\infty$  at an appropriate rate.

Estimators of both the mixing proportions and the component distributions follow readily along the lines of the proof of Theorem 1; see below. Since their asymptotic distribution will be driven by the large-sample behavior of the estimators of the quantities in (2.3), we start by deriving the statistical properties of these estimators.

### 2.1. Asymptotic theory for intermediate quantiles

Throughout this section we fix disjoint sets  $A, B$  and consider the asymptotic behavior of the estimators in (2.3).

Consistency only requires the following rate conditions.

ASSUMPTION 5 (Order statistics).  $\iota_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  and  $\kappa_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  as  $n \uparrow +\infty$ .

THEOREM 2 (Consistency). *If Assumptions 1–5 hold,*

$$\zeta_n^-(A, B) \xrightarrow{P} \zeta^-(A, B), \quad \zeta_n^+(A, B) \xrightarrow{P} \zeta^+(A, B),$$

as  $n \uparrow +\infty$ .

**Proof.** We prove the theorem for  $\zeta_n^+$ ; the proof for  $\zeta_n^-$  follows in a similar fashion. Write

$$\zeta_n^+ - \zeta^+ = (\zeta_n^+ - \zeta^{\kappa_{n_B}}) + (\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.4)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . For the second right-hand side term in (2.4) we have

$$\zeta^{\kappa_{n_B}} - \zeta^+ = \left( \frac{\lambda(A) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(A))}{\lambda(B) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(B))} - \frac{\lambda(A)}{\lambda(B)} \right) = O_p\left(\frac{1-H(r_{n_B})}{1-G(r_{n_B})}\right) = o_p(1),$$

by Assumptions 3(ii) and 5. To deal with the first right-hand side term in (2.4), recall that

$$\zeta_n^+ - \zeta^{\kappa_{n_B}} = \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)} - \frac{1 - F(r_{n_B}|A)}{1 - F(r_{n_B}|B)}.$$

Letting  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$  we thus have that

$$\begin{aligned} \zeta_n^+ - \zeta^{\kappa_{n_B}} &= \frac{(1 - F(r_{n_B}|A))\mathbb{G}_n(r_{n_B}|B)/\sqrt{n_B} - (1 - F(r_{n_B}|B))\mathbb{G}_n(r_{n_B}|A)/\sqrt{n_A}}{(1 - F_n(r_{n_B}|B))(1 - F(r_{n_B}|B))} \\ &= \frac{\sqrt{n_B}}{\kappa_{n_B}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right) \\ &= O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right), \end{aligned}$$

where the second equality uses  $1 - F_n(r_{n_B}|B) = \kappa_{n_B}/n_B$  and the last one follows by the law of the iterated logarithm for empirical processes. Thus, from Assumption 5 it follows that  $|\zeta_n^+ - \zeta^{\kappa_{n_B}}| = o_p(1)$ . This completes the proof.  $\square$

Deriving the limit distribution requires some more care, and three more assumptions. We first impose the following regularity condition on the component distributions.

ASSUMPTION 6. *G and H are absolutely continuous on  $\mathbb{R}$ .*

This assumption is very weak. Note that we do not require the existence of any moments of the component distributions. Thus, our results apply to heavy-tailed distributions such as Cauchy and Pareto distributions.

We will also complement Assumption 5 with an additional rate condition.

ASSUMPTION 7 (Order statistics (cont'd.)).  *$\iota_{n_B}/n_B \downarrow 0$  and  $\kappa_{n_B}/n_B \downarrow 0$  as  $n \uparrow +\infty$ .*

Where Assumption 5 required the order statistics to grow to ensure consistency, this assumption bounds this growth rate so that appropriately scaled versions of  $\zeta_n^+$  and  $\zeta_n^-$  have a limit distribution.

Finally, we will use an additional condition on the relative tails of the component distributions.

ASSUMPTION 8 (Tail rates).

- (i)  $G(\ell_{n_B})/H(\ell_{n_B}) = o_p(1/\sqrt{\iota_{n_B}})$ ; and
- (ii)  $(1 - H(r_{n_B}))/ (1 - G(r_{n_B})) = o_p(1/\sqrt{\kappa_{n_B}})$ .



Assumption 8 rules out distributions whose tails vanish too quickly and ensures that the limit distributions are free of asymptotic bias. We comment on Assumption 8 after we derive the limit distributions of our estimators.

Let  $\rho_{A,B} \equiv \mathbb{P}(X \in B)/\mathbb{P}(X \in A)$ . Note that  $0 < \rho_{A,B} < +\infty$  because of random sampling. Introduce

$$\begin{aligned}\sigma_-^2(A, B) &\equiv \zeta^-(A, B)^2 + \rho_{A,B} \zeta^-(A, B), \\ \sigma_+^2(A, B) &\equiv \zeta^+(A, B)^2 + \rho_{A,B} \zeta^+(A, B),\end{aligned}$$

Theorem 2 provides the asymptotic properties of the estimators in (2.2) and is the main building block for our subsequent results.

**THEOREM 3** (Asymptotic normality). *If Assumptions 1–8 hold, then as  $n \uparrow +\infty$ ,*

$$\begin{aligned}\sqrt{\iota_{n_B}}(\zeta_n^-(A, B) - \zeta^-(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, B)), \\ \sqrt{\kappa_{n_B}}(\zeta_n^+(A, B) - \zeta^+(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, B));\end{aligned}$$

*and these two estimators are asymptotically independent.*

**Proof.** We focus on the limit behavior of  $\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+)$  here; the proof of the result for  $\sqrt{\iota_n}(\zeta_n^- - \zeta^-)$  follows similar lines.

As in the proof of Theorem 2, write

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+) = \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) + \sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.5)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . Assumption 8 implies that

$$\sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+) = \sqrt{\kappa_{n_B}} O_p\left(\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})}\right) = o_p(1).$$

Hence, the second right-hand side term in (2.5) is asymptotically negligible.

We now turn to the first term in (2.5). From the proof of Theorem 2 we have that

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) = \sqrt{\frac{n_B}{\kappa_{n_B}}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right),$$

where  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$ . Let  $\alpha_n(u) \equiv \sqrt{n}(\mathcal{U}_n(u) - u)$  for  $\mathcal{U}_n$  the empirical cumulative distribution of an i.i.d. sample of size  $n$  from a uniform distribution on  $[0, 1]$ . By Assumption 6,  $F(y|S)$  is continuous in  $y$  for all  $S \subset \mathcal{X}$ . Therefore

$$\mathbb{G}_n(y|A) = \alpha_{n_A}(1 - F(y|A)) \text{ and } \mathbb{G}_n(y|B) = \alpha_{n_B}(1 - F(y|B))$$

by an application of the probability integral transform. Hence, we may write

$$\begin{aligned} \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) &= \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \\ &\quad - \sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A}(1 - F(r_{n_B}|A)). \end{aligned} \quad (2.6)$$

We study the asymptotic behavior of each of the right-hand side terms in turn.

Start with the first right-hand side term in (2.6). From the definition of the order statistic  $r_{n_B}$ , we find by adding and subtracting  $F_n(r_{n_B}|B)$  that

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} \left( 1 + \frac{\sqrt{n_B}}{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) \right);$$

or, defining  $\varepsilon_n \equiv -\sqrt{n_B}/\kappa_{n_B} \mathbb{G}_n(r_{n_B}|B)$ ,

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Therefore we can write

$$\zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) = \sqrt{2} \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right). \quad (2.7)$$

By the law of the iterated logarithm together with Assumption 5,

$$\varepsilon_n = -\frac{\sqrt{n_B}}{\kappa_{n_B}} O_{a.s.} \left( \sqrt{\ln \ln n_B} \right) = O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right) = o_{a.s.}(1).$$

Hence,  $(1 - \varepsilon_n)/2$  converges almost surely to  $1/2$ , and  $(1 - \varepsilon_n)/2 \in (0, 1)$  for  $n$  large enough. We may then apply Theorem 2.1 in Einmahl (1992) to establish the convergence in distribution of  $\sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right)$  to a normal random variable with mean zero and variance  $1/2$ . This, together with Equation (2.7) and an application of Slutsky's theorem, then implies that

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \zeta^{\kappa_{n_B}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \xrightarrow{d} \zeta^+ Z_B^+, \quad (2.8)$$

where  $Z_B^+$  is a standard normal random variable.

Now turn to the second right-hand side term in (2.6). First observe that

$$1 - F(r_{n_B}|A) = \zeta^{\kappa_{n_B}} (1 - F(r_{n_B}|B)) = \zeta^{\kappa_{n_B}} \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Using  $\rho_{A,B} = \lim_{n \uparrow +\infty} n_B/n_A$ , this gives

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) = \sqrt{2\rho_{A,B} \zeta^+} \sqrt{\frac{n_A}{2\tilde{\kappa}_{n_A}}} \alpha_{n_A} \left( \frac{2\tilde{\kappa}_{n_A}}{n_A} \frac{1 - \varepsilon_n}{2} \right) + o_p(1),$$

where  $\tilde{\kappa}_{n_A} \equiv (\kappa_{n_B} \zeta^{\kappa_{n_B}})/(n_B/n_A)$ . As  $\tilde{\kappa}_{n_A}$  satisfies Assumption (1.5) of Theorem 2.1 in Einmahl (1992) we may apply his theorem again to obtain

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) \xrightarrow{d} \sqrt{\rho_{A,B} \zeta^+} Z_A^+, \quad (2.9)$$

where  $Z_A^+$  is a standard-normal random variable which, by random sampling, is independent of  $Z_B^+$ .

Combining (2.6) with (2.8) and (2.9) then gives

$$\sqrt{\kappa_{n_B}} (\zeta_n^+ - \zeta^{\kappa_{n_B}}) \xrightarrow{d} \zeta^+ Z_B^+ - \sqrt{\rho_{A,B} \zeta^+} Z_A^+,$$

as claimed. This concludes the proof.  $\square$

We finish this section with two examples that specialize Assumption 8 to densities with log-concave tails and Pareto tails, respectively. In both cases, Assumption 8 is implied by the rate conditions in Assumption 7.

EXAMPLE 6 (Log-concave tails). *Suppose that  $G$  and  $H$  have log-concave tails; and for notational simplicity, assume that*

$$-\ln(1 - G(y)) \sim \left( \frac{y}{\sigma_G^+} \right)^{\alpha_G^+}, \quad -\ln(1 - H(y)) \sim \left( \frac{y}{\sigma_H^+} \right)^{\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for real numbers  $\alpha_G^+, \alpha_H^+ > 1$  and  $\sigma_G^+, \sigma_H^+ > 0$ , and

$$-\ln G(y) \sim \left( \frac{-y}{\sigma_G^-} \right)^{\alpha_G^-}, \quad -\ln H(y) \sim \left( \frac{-y}{\sigma_H^-} \right)^{\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

for real numbers  $\alpha_G^-, \alpha_H^- > 1$  and  $\sigma_G^-, \sigma_H^- > 0$ . Then Assumption 7 implies Assumption 8 if both

(i)  $\alpha_G^+ < \alpha_H^+$ , or  $\alpha_G^+ = \alpha_H^+$  and  $\sigma_G^+ > \sigma_H^+$ ; and

(ii)  $\alpha_G^- > \alpha_H^-$ , or  $\alpha_G^- = \alpha_H^-$  and  $\sigma_G^- < \sigma_H^-$

hold.

**Proof.** We verify the second rate; the first follows similarly. Throughout, fix the set  $B$ . Assumptions 3(ii) and 7 imply that

$$\begin{aligned} 1 - F(r_{n_B}|B) &= (1 - G(r_{n_B})) \lambda(B) + (1 - H(r_{n_B})) (1 - \lambda(B)) \\ &= (1 - G(r_{n_B})) (\lambda(B) + o_p(1)). \end{aligned}$$

Further, because  $\kappa_{n_B}/n_B = 1 - F_n(r_{n_B}|B)$ , adding and subtracting  $F(r_{n_B}|B)$  gives

$$\begin{aligned} \frac{\kappa_{n_B}}{n_B} &= (1 - F(r_{n_B}|B)) + (F_n(r_{n_B}|B) - F(r_{n_B}|B)) \\ &= (1 - F(r_{n_B}|B)) + O_{a.s.}(\sqrt{(\ln \ln n_B)/n_B}). \end{aligned}$$

Because  $(\ln \ln n_B)/n_B \rightarrow 0$ , put together, we find

$$\frac{\kappa_{n_B}}{n_B} = C (1 - G(r_{n_B})) (1 + o_p(1))$$

for some constant  $C$ . With  $G$  and  $H$  having log-concave tails, it then follows from this expression that, asymptotically,  $r_{n_B}$  behaves like  $\alpha_G^+ \sqrt[n_B]{\ln n_B}$ . So, because

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} \sim \exp \left\{ \left( \frac{r_{n_B}}{\sigma_G^+} \right)^{\alpha_G^+} - \left( \frac{r_{n_B}}{\sigma_H^+} \right)^{\alpha_H^+} \right\},$$

we have that

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} = \begin{cases} O_p \left( \exp(-(\ln n_B)^{\alpha_H^+/\alpha_G^+}) \right) & \text{if } \alpha_H^+ > \alpha_G^+ \\ O_p(1/n_B) & \text{if } \alpha_H^+ = \alpha_G^+ \text{ and } \sigma_H^+ < \sigma_G^+ \end{cases},$$

from which the conclusion follows.  $\square$

Example 6 does not cover location models with log-concave distributions for which the  $\alpha$  and  $\sigma$  parameters of  $H$  equal those of  $G$ . This includes the location model with Gaussian errors, for which  $\alpha = 2$  and  $\sigma$  is the common standard error. While our estimator remains

consistent in such cases, we do not know of general results on tail empirical processes that would yield the asymptotic distribution of the estimator in this knife-edge case. To assess the extent to which the failure of Assumption 8 may play a role for inference, our simulation experiments in Section 4 include a Gaussian location model.

EXAMPLE 7 (Pareto tails). *Let  $C$  denote a generic constant. Suppose that  $G$  and  $H$  have Pareto tails, i.e.,*

$$(1 - G(y)) \sim C y^{-\alpha_G^+}, \quad (1 - H(y)) \sim C y^{-\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

*for positive real numbers  $\alpha_H^+ > \alpha_G^+$  and*

$$G(y) \sim C (-y)^{-\alpha_G^-}, \quad H(y) \sim C (-y)^{-\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

*for positive real numbers  $\alpha_G^- < \alpha_H^-$ . Then Assumption 7 implies Assumption 8.*

**Proof.** The argument is very similar to the one that was used to verify Example 6. We treat the right tail only. We have

$$\frac{\kappa_{n_B}}{n_B} = (1 - G(r_{n_B})) (1 + o_p(1)) = C r^{-\alpha_G^+} (1 + o_p(1)).$$

Assumption 8 requires that  $(1 - H(r_{n_B})) / (1 - G(r_{n_B})) = o(1/\sqrt{\kappa_n})$ , that is, that  $r_{n_B}^{\alpha_G^+ - \alpha_H^+} = o_p(1/\sqrt{\kappa_{n_B}})$ . This rate condition is satisfied when

$$\left( \frac{n_B}{\kappa_{n_B}} \right)^{\frac{\alpha_G^+ - \alpha_H^+}{\alpha_G^+}} = o_p \left( \frac{1}{\sqrt{\kappa_{n_B}}} \right),$$

which can be achieved by setting  $\kappa_{n_B} = o(n_B^{\gamma^+})$  for

$$\gamma^+ \equiv \frac{\alpha_H^+ - \alpha_G^+}{\alpha_H^+ - \alpha_G^+/2}. \quad (2.10)$$

This condition is weaker than Assumption 7 and is therefore implied by it.  $\square$

Example 7 shows that our methods are well suited to deal with Pareto tails. Pareto tails show up in many economic applications. A time-honored example is income and wealth distributions (Atkinson et al. 2011), which are often modeled as a log-normal for

most quantiles, combined with a Pareto right tail. More generally, “power laws” have become a popular tool in finance, in studies of firm growth, and in urban economics (see [Gabaix 2009](#) for a recent survey, and [Acemoglu et al. 2012](#) for an application to business cycles.) Many recent models of monopolistic competition, as used in international trade for instance, also assume that productivities are Pareto-distributed ([Arkolakis et al. 2012](#)).

Identification only requires that the tail index of  $H$  be larger than that of  $G$ , that is,  $\alpha_H > \alpha_G$ . Denote  $c = \alpha_H/\alpha_G > 1$ . Equation (2.10) then gives a convergence rate arbitrarily close to  $n^{-\beta/2}$  for  $\beta = 2(c-1)/(2c-1)$ . For example, if  $c = 2$  then  $\beta = 2/3$  and our estimators will converge slightly more slowly than  $n^{-1/3}$ . However, as  $c$  increases,  $\beta$  rapidly converges to one, and our estimators will converge at close to the  $n^{-1/2}$  parametric rate.

## 2.2. Mixing proportions

Fix  $x \in \mathcal{X}$  and consider estimating  $\lambda(x)$ . Set  $A = \mathcal{X} - x$  and  $B = x$  in (2.2) and solve for  $\lambda(x)$  to get

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}.$$

The mixing proportion  $\lambda$  need not be a strictly monotonic function. Estimating  $\lambda(x)$  by an average of plug-in estimates of (1.3) could therefore be problematic, as the denominator in (1.3) can be zero or be arbitrarily close to it for some pairs of values  $(x', x'')$ .

We instead estimate the mixing proportion at  $X = x$  by a plug-in estimator based on (2.3), that is,

$$\lambda_n(x) \equiv \frac{1 - \zeta_n^-(A, x)}{\zeta_n^+(A, x) - \zeta_n^-(A, x)}.$$

This estimator uses observations with  $X_i \neq x$  in a way that immunizes it against small or zero denominators.

To present the asymptotic variance of this estimator we need to define

$$\begin{aligned} d^-(x) &\equiv \frac{1 - \zeta^+(A, x)}{(\zeta^+(A, x) - \zeta^-(A, x))^2}, \\ d^+(x) &\equiv \frac{\zeta^-(A, x) - 1}{(\zeta^+(A, x) - \zeta^-(A, x))^2}. \end{aligned} \tag{2.11}$$

The speed of convergence and the asymptotic distribution of the  $\lambda_n(x)$  depend on the ratio  $c_x \equiv \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

**THEOREM 4** (Mixing proportions). *Under the conditions of Theorem 2,*

$$|\lambda_n(x) - \lambda(x)| = o_p(1)$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, d^-(x)^2 \sigma_-^2(A, x) + c_x d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x < +\infty, \\ \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, c_x^{-1} d^-(x)^2 \sigma_-^2(A, x) + d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x > 0, \end{aligned}$$

as  $n \uparrow +\infty$ .

**Proof.** The consistency claim follows directly from Theorem 2 by an application of the continuous mapping theorem.

To establish the asymptotic distribution, note that Theorem 2 states that

$$\begin{aligned} \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, x)), \\ \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, x)), \end{aligned}$$

and that  $\zeta_n^-(x)$  and  $\zeta_n^+(x)$  are asymptotically independent. An expansion around  $\zeta^-(A, x)$  and  $\zeta^+(A, x)$  then yields

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) \\ &\quad + d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) \sqrt{\frac{\iota_{n_x}}{\kappa_{n_x}}} + o_p(1), \end{aligned}$$

which has the limit distribution stated in the theorem if  $c_x$  is finite. Also, by the same argument,

$$\begin{aligned} \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(x) - \zeta^+(x)) \\ &\quad + d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(x) - \zeta^-(x)) \sqrt{\frac{\kappa_{n_x}}{\iota_{n_x}}} + o_p(1) \end{aligned}$$

converges in distribution as stated in the theorem if  $c_x$  is non-zero. This verifies the claims and proves the theorem.  $\square$

### 2.3. Component distributions

To estimate the component distributions let  $A, B$  be two sets that partition  $\mathcal{X}$ . Equations (1.5) and (1.6) then suggest the estimators

$$\begin{aligned} H_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^+(B, A)} (F_n(y|A) - F_n(y|B)), \\ G_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)). \end{aligned} \quad (2.12)$$

Fix  $A, B$  and omit them as arguments, so  $G_n(y) = G_n(y; A, B)$  and  $H_n(y) = H_n(y; A, B)$ .

To state their asymptotic behavior, let

$$\begin{aligned} d_G(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^-(B, A))^2}, \\ d_H(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^+(B, A))^2}, \end{aligned}$$

and let  $\|\cdot\|_\infty$  denote the supremum norm.

**THEOREM 5.** *Under the conditions of Theorem 2,*

$$\|G_n - G\|_\infty = o_p(1), \quad \|H_n - H\|_\infty = o_p(1),$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_A}}(G_n(y) - G(y)) &\xrightarrow{d} \mathcal{N}(0, d_G(A, B; y)^2 \sigma_-^2(B, A)), \\ \sqrt{\kappa_{n_A}}(H_n(y) - H(y)) &\xrightarrow{d} \mathcal{N}(0, d_H(A, B; y)^2 \sigma_+^2(B, A)), \end{aligned}$$

as  $n \uparrow +\infty$  for each  $y \in \mathbb{R}_+$ .

**Proof.** Consistency follows by Theorem 2 and the Glivenko-Cantelli theorem.

We establish the asymptotic distribution of  $G_n$ ; the result for  $H_n$  follows by the same argument.

First note that

$$\sqrt{\iota_{n_A}}(G_n(y) - G(y)) = T_1 + T_2 + T_3$$



for

$$\begin{aligned} T_1 &\equiv \sqrt{\iota_{n_A}}(F_n(y|A) - F(y|A)), \\ T_2 &\equiv -\frac{1}{1 - \zeta^-(B, A)} \sqrt{\iota_{n_A}} \left( \{F_n(y|A) - F(y|A)\} - \{F_n(y|B) - F(y|B)\} \right), \\ T_3 &\equiv -(F_n(y|A) - F_n(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right). \end{aligned}$$

By the Glivenko-Cantelli theorem,  $T_1 = o_p(1)$  and  $T_2 = o_p(1)$  while

$$T_3 = -(F(y|A) - F(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right) + o_p(1).$$

A linearization of this expression in  $\zeta_n^-(B, A) - \zeta^-(B, A)$  together with an application of Theorem 2 to the partition  $A, B$  then yields the result.  $\square$

### 3. SPECIFICATION TESTING

The tail restriction in Assumption 3 and the exclusion restriction on  $X$  are the two key ingredients that lead to identification via Theorem 1. The proof of the theorem shows that both restrictions have testable implications.

#### 3.1. Testing for tail dominance

As is evident from the proof of Theorem 1, Assumption 3 implies that

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}$$

for any  $A \subset \mathcal{X} - x$ . Denote  $\lambda_n(x; A)$  the corresponding plug-in estimator of  $\lambda(x)$ . Any such estimator will be consistent and asymptotically-normal under the conditions of Theorem 2 and Theorem 3, respectively. Therefore, letting  $A, B$  be two disjoint subsets of  $\mathcal{X} - x$ , a simple specification test for the tail restriction amounts to testing whether the difference

$$\lambda_n(x; A) - \lambda_n(x; B)$$

is statistically different from zero.

As before, the limit distribution of this difference depends on  $c_x = \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

We define

$$\begin{aligned} \Sigma_- &\equiv d^-(x, A) \left\{ d^-(x, A) \sigma_-^2(A, x) - d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) \right\} \\ &\quad + d^-(x, B) \left\{ d^-(x, B) \sigma_-^2(B, x) - d^-(x, A) \zeta^-(A, x) \zeta^-(B, x) \right\}, \end{aligned}$$

and

$$\begin{aligned} \Sigma_+ \equiv & d^+(x, A) \left\{ d^+(x, A) \sigma_+^2(A, x) - d^+(x, B) \zeta^+(A, x) \zeta^+(B, x) \right\} \\ & + d^+(x, B) \left\{ d^+(x, B) \sigma_+^2(B, x) - d^+(x, A) \zeta^+(A, x) \zeta^+(B, x) \right\}, \end{aligned}$$

in Theorem 6.

THEOREM 6 (A test for tail dominance). *Under the conditions of Theorem 3,*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{\Sigma_- + c_x \Sigma_+ / \sqrt{\ell_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x < +\infty$ , while

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{c_x^{-1} \Sigma_- + \Sigma_+ / \sqrt{\kappa_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x > 0$ , where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** Fix the sets  $A, B$ . We treat only the case  $c_x > 0$ . Our assumptions imply the joint limit result

$$\sqrt{\kappa_{n_x}} \begin{pmatrix} \lambda_n(x; A) - \lambda(x) \\ \lambda_n(x; B) - \lambda(x) \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\kappa^{AA} & \Sigma_\kappa^{AB} \\ \Sigma_\kappa^{AB} & \Sigma_\kappa^{BB} \end{pmatrix} \right),$$

where

$$\begin{aligned} \Sigma_\kappa^{AA} &\equiv c_x^{-1} d^-(x, A)^2 \sigma_-^2(A, x) + d^+(x, A)^2 \sigma_+^2(A, x), \\ \Sigma_\kappa^{BB} &\equiv c_x^{-1} d^-(x, B)^2 \sigma_-^2(B, x) + d^+(x, B)^2 \sigma_+^2(B, x), \end{aligned}$$

follow directly from an application of Theorem 4, and

$$\Sigma_\kappa^{AB} \equiv c_x^{-1} d^-(x, A) d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) + d^+(x, A) d^+(x, B) \zeta^+(A, x) \zeta^+(B, x)$$

follows readily from the asymptotic-linearity statement at the end of the proof of Theorem 3. It follows that the transformation  $\sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda_n(x))$  is asymptotically normal with variance

$$\Sigma_\kappa^{AA} + \Sigma_\kappa^{BB} - 2 \Sigma_\kappa^{AB} = c_x^{-1} \Sigma_- + \Sigma_+.$$

This proves the result. □

### 3.2. Testing the exclusion restriction

An implication of the exclusion restriction in (1.1) is that the estimators of the component distributions  $G$  and  $H$  in (2.12), when based on different subsets of  $\mathcal{X}$ , should co-incide with one another, up to sampling error.

Theorem 7 provides the relevant asymptotic variance to perform this test. In it we use

$$\begin{aligned}\Sigma_G = & d_G(A, C; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\} \\ & + d_G(A, B; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\}\end{aligned}$$

and

$$\begin{aligned}\Sigma_H = & d_H(A, C; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\} \\ & + d_H(A, B; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\},\end{aligned}$$

where the triple  $A, B, C$  constitutes any partition of  $\mathcal{X}$ .

**THEOREM 7** (A test of the exclusion restriction). *Under the conditions of Theorem 3*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{G_n(y; A, B) - G_n(y; A, C)}{\sqrt{\Sigma_G}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

and

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{H_n(y; A, B) - H_n(y; A, C)}{\sqrt{\Sigma_H}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** We consider only the case of  $G$ . The difference  $G_n(y; A, B) - G_n(y; A, C)$  equals

$$\frac{1}{1 - \zeta_n^-(C, A)} (F_n(y|A) - F_n(y|C)) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)).$$

Proceeding as in the proof to Theorem 6 immediately yields the asymptotic-distribution result that implies the theorem.  $\square$

## 4. SIMULATION EXPERIMENTS

In our numerical illustrations we will work with the family of skew-normal distributions (Azzalini 1985.) The skew-normal distribution with location  $\mu$ , positive scale  $\sigma$ , and

skewness parameter  $\beta$  multiplies the density of  $\mathcal{N}(\mu, \sigma^2)$  by a term that skews it to the right if  $\beta > 0$  and to the left if  $\beta < 0$ . Moreover,

$$f(x; \mu, \sigma, \beta) \equiv \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \times \frac{\Phi\left(\beta \frac{x - \mu}{\sigma}\right)}{\Phi(0)}.$$

Its mean and variance are  $\mu + \sigma\delta\sqrt{\frac{2}{\pi}}$  and  $\sigma^2\left(1 - \frac{2\delta^2}{\pi}\right)$ , respectively, where  $\delta \equiv \beta/\sqrt{1 + \beta^2}$ . Clearly,

$$f(x; \mu, \sigma, \beta) \rightarrow \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)$$

as  $\beta \rightarrow 0$ .

In our simulations we will consider data generating processes where the outcome is generated as

$$Y = T V_G + (1 - T) V_H, \quad (4.1)$$

where  $T$  is a latent binary variable, and  $V_G \sim G$  and  $V_H \sim H$ . Both error distributions  $G$  and  $H$  are skewed-normal distributions with parameters  $\mu_G, \sigma_G, \beta_G$  and  $\mu_H, \sigma_H, \beta_H$ , respectively.

From [Capitanio \(2010\)](#) it follows that Assumption 8 holds if  $G$  is right-skewed and  $H$  is left-skewed. We will consider designs where  $\beta_G > 0$  and  $\beta_H < 0$  to verify our asymptotics.

When  $\beta_G = \beta_H = 0$ , (4.1) collapses to a standard location model with normal errors

$$Y = (\mu_G - \mu_H) T + V, \quad V \sim \mathcal{N}(0, \sigma_G^2 + \sigma_H^2). \quad (4.2)$$

The identifying tail condition in Assumption 3 still holds if  $\mu_G > \mu_H$ , and our estimators remain consistent. However, Assumption 8 now fails and so we may expect poor inference in this design.

In our experiments we generate a binary  $X$  with  $\mathbb{P}(X = 1) = \frac{1}{2}$  and fix conditional probabilities as

$$\begin{aligned} \mathbb{P}(T = 0|X = 0) &= \frac{3}{4}, & \mathbb{P}(T = 1|X = 0) &= \frac{1}{4}, \\ \mathbb{P}(T = 1|X = 1) &= \frac{1}{4}, & \mathbb{P}(T = 0|X = 1) &= \frac{3}{4}. \end{aligned}$$

We present results for data generating processes where  $\mu_G = \mu = -\mu_H$  and  $\beta_G = \beta = -\beta_H$ . We use the designs  $\mu = 0$  and  $\beta \in \{2.5, 5\}$  to evaluate the adequacy of our

asymptotic arguments for small-sample inference. We also look at the performance of our estimators when  $\mu \in \{.5, 1\}$  and  $\beta = 0$ , which yields the Gaussian location model in (4.2). We fix  $\sigma_G = \sigma_H = 1$  throughout. For each of these designs we consider choices of the empirical quantiles as

$$\iota_{n_x} = C(n_x \ln \ln n_x)^{6/10}, \quad \kappa_{n_x} = C(n_x \ln \ln n_x)^{6/10},$$

for several choices of the constant  $C$ . All of these choices are in line with our asymptotic arguments. The larger the constant  $C$  the more conservative the choice of intermediate quantile,

$$q_\ell \equiv \frac{\iota_{n_x}}{n_x}, \quad q_r \equiv \frac{n_x - \kappa_{n_x}}{n_x},$$

for a given sample size.

We run experiments for sample sizes  $n \in \{500; 1,000, 2,500; 5,000; 10,000; 25,000\}$ . We report (the average over the replications of)  $q_\ell$  and  $q_r$  along with the estimation results to get an idea of how far in the tails of the component distributions we are going to obtain the results. A data-driven determination of the constant  $C$  is challenging and is left for future research. All tables and figures concerning our Monte Carlo experiment are collected at the end of the paper.

Tables 1–4 report the simulation results for the mixing proportions  $\lambda(0)$  and  $\lambda(1)$  for the data generating processes with  $\mu = 0, \beta = 5$ ;  $\mu = 0, \beta = 2.5$ ;  $\mu = 1, \beta = 0$ ; and  $\mu = .5, \beta = 0$ , respectively. Each table contains the bias, standard deviation (SD), ratio of the (average over the replications of the) estimated standard error to the standard deviation (SE/SD), and the coverage of 95% confidence intervals (CI95). All these statistics were computed from 10,000 Monte Carlo replications.

The results in Table 1 and Table 2 support our asymptotic theory. For each design and all choices of the tuning parameter  $C$ , the bias and standard deviation shrink to zero as  $n \uparrow +\infty$ ; and the bias is small relative to the standard error. Furthermore,  $\text{SE}/\text{SD} \rightarrow 1$  and the coverage rates of the confidence intervals are close to .95 in large samples. The variability of the point estimates is somewhat overestimated when  $n$  is very small and  $C$  is chosen conservatively. Together with the relatively small bias, this implies that confidence intervals are slightly conservative. For  $C = .5$ , coverage rates are close to .95,

even for the smallest samples considered, and for all  $C$ , the coverage rates move fairly quickly toward .95 as  $n$  increases.

Tables 3 and 4 present the results for the pure location model with Gaussian errors, where the tail conditions of Assumption 8 fail. The difference between the two designs is the distance between the component distributions. For Table 3,  $G$  is centered at 1 while  $H$  is centered at  $-1$ , so that  $\mu_G - \mu_H = 2$ . For Table 4,  $G$  and  $H$  are closer to each other:  $\mu_G - \mu_H = 1$ . In the first of these designs the bias in the point estimates is somewhat larger than in the skewed designs. Nonetheless, the bias is still small relative to the standard deviation. Furthermore, the coverage of the confidence intervals displays a similar pattern as in Tables 1 and 2, and is excellent for all values of  $C$  when  $n$  is not too small. When we move to Table 4 the bias increases further. The bias still shrinks to zero as  $n$  grows, confirming that our estimator remains consistent. However, the bias is not negligible relative to the standard deviation; the coverage of the confidence intervals deteriorates as  $n$  grows, and inference becomes unreliable.

We next turn to the results for the component distributions. For clarity we present the results by means of a series of plots. Figures 1–4 contain the results corresponding to Tables 1–4, respectively. To save on space we only present results for one value of  $C$ —the effect of larger values of  $C$  on  $G_n$  and  $H_n$  is similar to that on  $\lambda_n$  documented before. Each figure presents results for  $G_n$  (left plots) and  $H_n$  (right plots) for (from top to bottom)  $n = 1,000$ ,  $n = 2,500$ ,  $n = 5,000$ , and  $n = 10,000$ . Each plot contains the mean of the point estimates and the mean of 95% confidence bounds constructed using a plug-in estimator of the asymptotic variance in Theorem 5 (dashed lines). Each plot also contains the true component distribution and 95% confidence bounds constructed using the empirical standard deviation over the Monte Carlo replications (solid grey lines). We vary the range of the vertical axis across the plots in a given figure to enhance visibility.

Figures 1 and 2 again confirm our asymptotics. Indeed, moving down the rows,  $G_n$  and  $H_n$  converge to  $G$  and  $H$  respectively, while the estimated standard error converges to the Monte Carlo standard deviation. The bias in the point estimates is small across all plots. For small  $n$  the estimated standard errors are somewhat too small; this underestimation is more severe for  $H_n$  than for  $G_n$ .

Figures 3 and 4 concern the Gaussian location model. The results are in line with

our findings concerning the mixing proportions. In the design where  $\mu_G - \mu_H = 2$  our estimators do well in spite of Assumption 8 not holding. In Figure 4, however, the asymptotic bias in  $G_n$  and  $H_n$  becomes visible. While the variability of the point estimates is correctly captured by our asymptotic-variance estimator, the confidence bounds settle around an incorrect curve.

### CONCLUDING REMARKS

We conducted most of our analysis with a mixture of two components. However, some of our results would extend to a version of (1.1) with a larger number of components. Suppose that the mixture has  $J$  irreducible components, as in

$$F(y|x) = \sum_{j=1}^J \lambda_j(x) G_j(y),$$

in obvious notation. Henry et al. (2014) showed that the mixture components and mixing proportions are only identified up to  $J(J-1)$  inequality-constrained real parameters in general.

Tail dominance restrictions can still be quite powerful. Take  $J = 3$  for instance, and assume that  $G_1$  dominates in the left tail and  $G_3$  dominates in the right tail. Then it is easy to adapt the proof of Theorem 1 to prove that the behavior of  $F(y|x)$  in the left tail identifies the function  $\lambda_1$  up to a multiplicative constant, and that the behavior of  $F(y|x)$  in the right tail identifies the function  $\lambda_3$  up to another multiplicative constant. Imposing the values of the mixing proportions at one particular value of  $x$  would be enough to point identify all elements of the model, for instance; and it would be easy to adapt our estimators and tests to such a setting. Whether such additional restrictions are plausible is, of course, highly model-dependent.

**Table 1.**  $\mu = 0$  and  $\beta = 5$ 

$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0115	−.0083	.0951	.0914	1.0790	1.0384	.9714	.9694
1,000	.059	.940	.0060	−.0059	.0693	.0701	1.0554	1.0392	.9688	.9682
2,500	.043	.957	.0038	−.0029	.0510	.0506	1.0321	1.0393	.9624	.9636
5,000	.034	.966	.0016	−.0023	.0406	.0407	1.0246	1.0246	.9588	.9592
10,000	.026	.974	.0012	−.0011	.0328	.0325	1.0106	1.0213	.9560	.9572
25,000	.018	.982	.0013	−.0007	.0245	.0244	1.0108	1.0149	.9526	.9568
$C = 1$										
500	.172	.848	.0044	−.0046	.0534	.0535	1.1937	1.1907	.9904	.9876
1,000	.120	.880	.0024	−.0035	.0439	.0446	1.1358	1.1220	.9764	.9752
2,500	.087	.914	.0012	−.0010	.0332	.0337	1.1031	1.0847	.9680	.9668
5,000	.067	.933	.0012	−.0005	.0270	.0271	1.0814	1.0726	.9664	.9670
10,000	.052	.947	.0007	−.0003	.0225	.0222	1.0360	1.0519	.9566	.9616
25,000	.037	.963	.0006	−.0005	.0167	.0167	1.0459	1.0460	.9594	.9600
$C = 1.5$										
500	.227	.773	.0198	−.0198	.0328	.0333	1.6715	1.6482	.9982	.9984
1,000	.179	.821	.0046	−.0037	.0316	.0315	1.2931	1.2933	.9944	.9920
2,500	.129	.870	.0013	.0000	.0257	.0256	1.1575	1.1579	.9774	.9762
5,000	.101	.898	.0010	−.0005	.0208	.0215	1.1427	1.1070	.9756	.9720
10,000	.078	.922	.0002	−.0010	.0175	.0174	1.0873	1.0962	.9646	.9710
25,000	.056	.944	.0004	.0000	.0136	.0136	1.0453	1.0474	.9632	.9604

**Table 2.**  $\mu = 0$  and  $\beta = 2.5$ 

	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0093	−.0122	.0941	.0934	1.0161	1.0315	.9718	.9780
1,000	.059	.940	.0066	−.0072	.0722	.0718	1.0151	1.0194	.9646	.9652
2,500	.043	.957	.0029	−.0040	.0512	.0519	1.0279	1.0212	.9610	.9612
5,000	.034	.966	.0016	−.0026	.0407	.0407	1.0248	1.0255	.9546	.9596
10,000	.026	.974	.0012	−.0015	.0323	.0326	1.0287	1.0193	.9548	.9626
25,000	.018	.982	.0007	−.0009	.0243	.0244	1.0182	1.0137	.9564	.9580
$C = 1$										
500	.172	.848	.0160	−.0174	.0519	.0518	1.2924	1.2969	.9954	.9972
1,000	.120	.880	.0069	−.0084	.0429	.0430	1.1865	1.1898	.9862	.9890
2,500	.087	.914	.0032	−.0024	.0334	.0338	1.1035	1.0900	.9716	.9684
5,000	.067	.933	.0018	−.0014	.0270	.0275	1.0868	1.0648	.9700	.9666
10,000	.052	.947	.0001	−.0010	.0220	.0219	1.0569	1.0661	.9636	.9652
25,000	.037	.963	.0005	−.0005	.0168	.0168	1.0376	1.0356	.9603	.9585
$C = 1.5$										
500	.227	.773	.0491	−.0496	.0361	.0365	1.7108	1.6925	.9908	.9902
1,000	.179	.821	.0245	−.0248	.0304	.0303	1.4562	1.4670	.9920	.9926
2,500	.129	.870	.0083	−.0080	.0248	.0250	1.2350	1.2249	.9868	.9872
5,000	.101	.898	.0030	−.0035	.0211	.0210	1.1386	1.1418	.9774	.9788
10,000	.078	.922	.0015	−.0016	.0172	.0172	1.1103	1.1090	.9722	.9720
25,000	.056	.944	.0005	−.0007	.0136	.0136	1.0496	1.0455	.9604	.9568

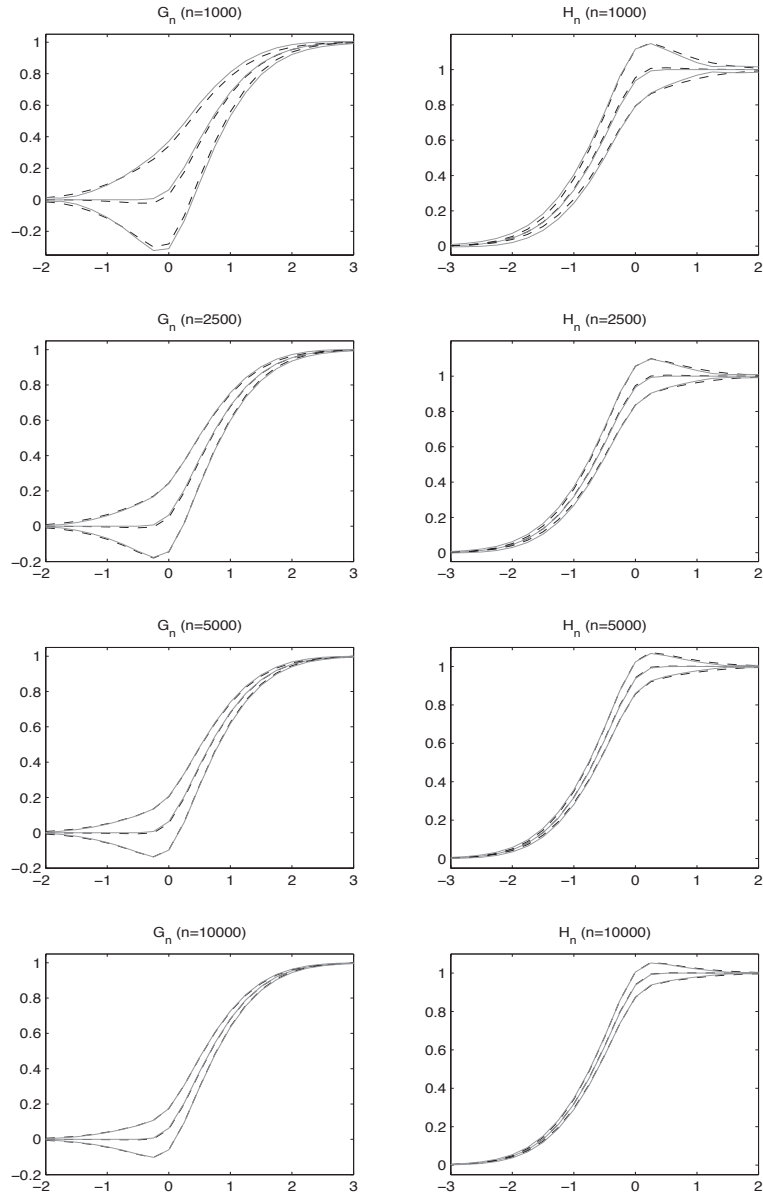


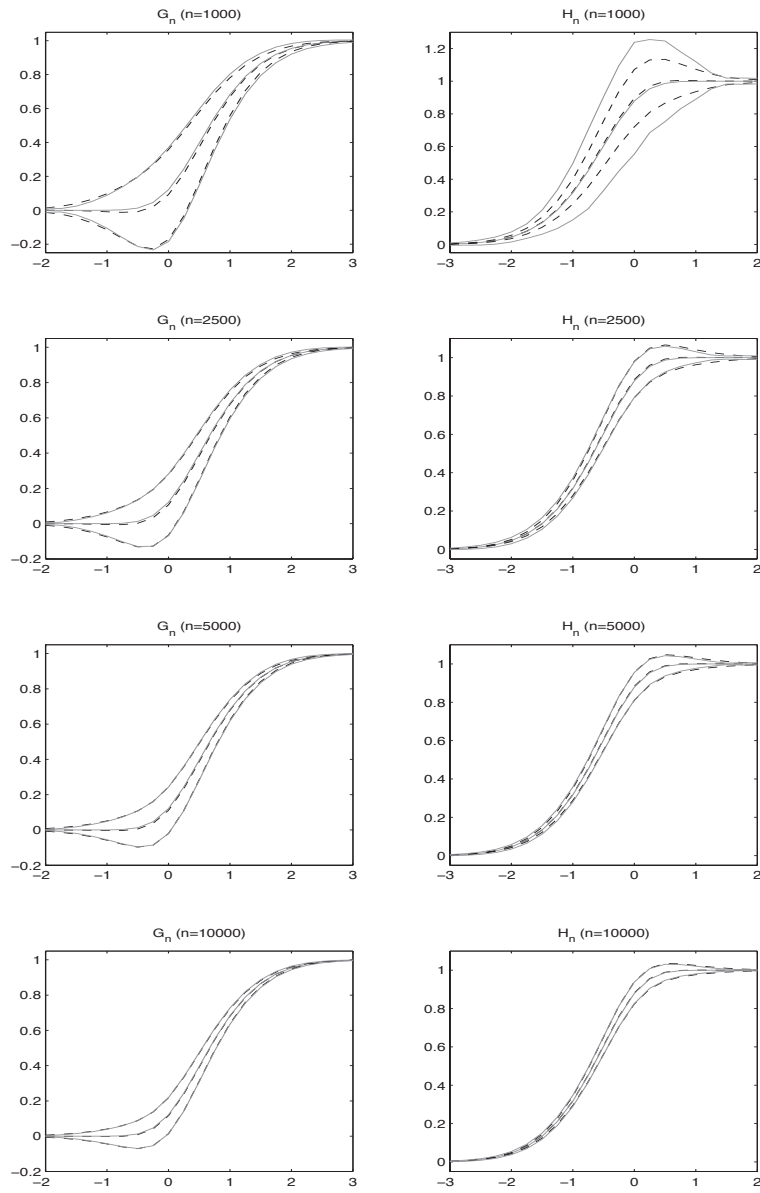
**Table 3.**  $\mu = 1$  and  $\beta = 0$ 

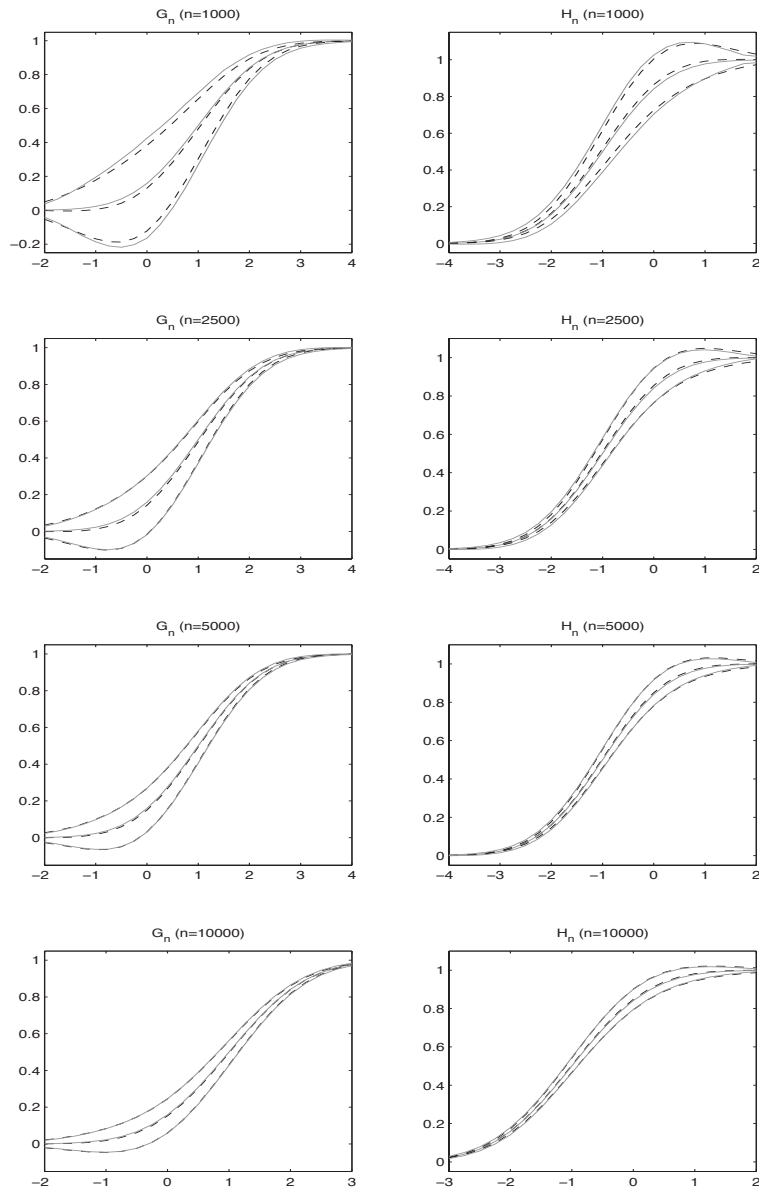
$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0220	−.0227	.0966	.0948	1.0479	1.0710	.9869	.9866
1,000	.059	.940	.0144	−.0164	.0720	.0728	1.0589	1.0518	.9807	.9810
2,500	.043	.957	.0092	−.0105	.0519	.0527	1.0476	1.0331	.9705	.9697
5,000	.034	.966	.0067	−.0069	.0411	.0415	1.0372	1.0283	.9673	.9651
10,000	.026	.974	.0050	−.0048	.0327	.0324	1.0344	1.0449	.9614	.9622
25,000	.018	.982	.0030	−.0033	.0246	.0245	1.0188	1.0246	.9576	.9586
$C = 1$										
500	.172	.848	.0385	−.0395	.0576	.0575	1.2942	1.2958	.9948	.9944
1,000	.120	.880	.0268	−.0267	.0459	.0453	1.2162	1.2304	.9867	.9869
2,500	.087	.914	.0154	−.0166	.0339	.0344	1.1554	1.1417	.9786	.9770
5,000	.067	.933	.0110	−.0120	.0274	.0278	1.1168	1.1036	.9722	.9691
10,000	.052	.947	.0079	−.0084	.0225	.0224	1.0752	1.0823	.9641	.9656
25,000	.037	.963	.0055	−.0055	.0169	.0173	1.0602	1.0361	.9603	.9562
$C = 1.5$										
500	.227	.773	.0718	−.0714	.0421	.0418	1.6251	1.6382	.9726	.9739
1,000	.179	.821	.0476	−.0474	.0343	.0340	1.4396	1.4479	.9619	.9632
2,500	.129	.870	.0277	−.0282	.0269	.0270	1.2510	1.2445	.9502	.9487
5,000	.101	.898	.0188	−.0191	.0217	.0219	1.1929	1.1831	.9510	.9473
10,000	.078	.922	.0133	−.0132	.0178	.0179	1.1330	1.1285	.9443	.9469
25,000	.056	.944	.0086	−.0084	.0135	.0136	1.0949	1.0895	.9424	.9458

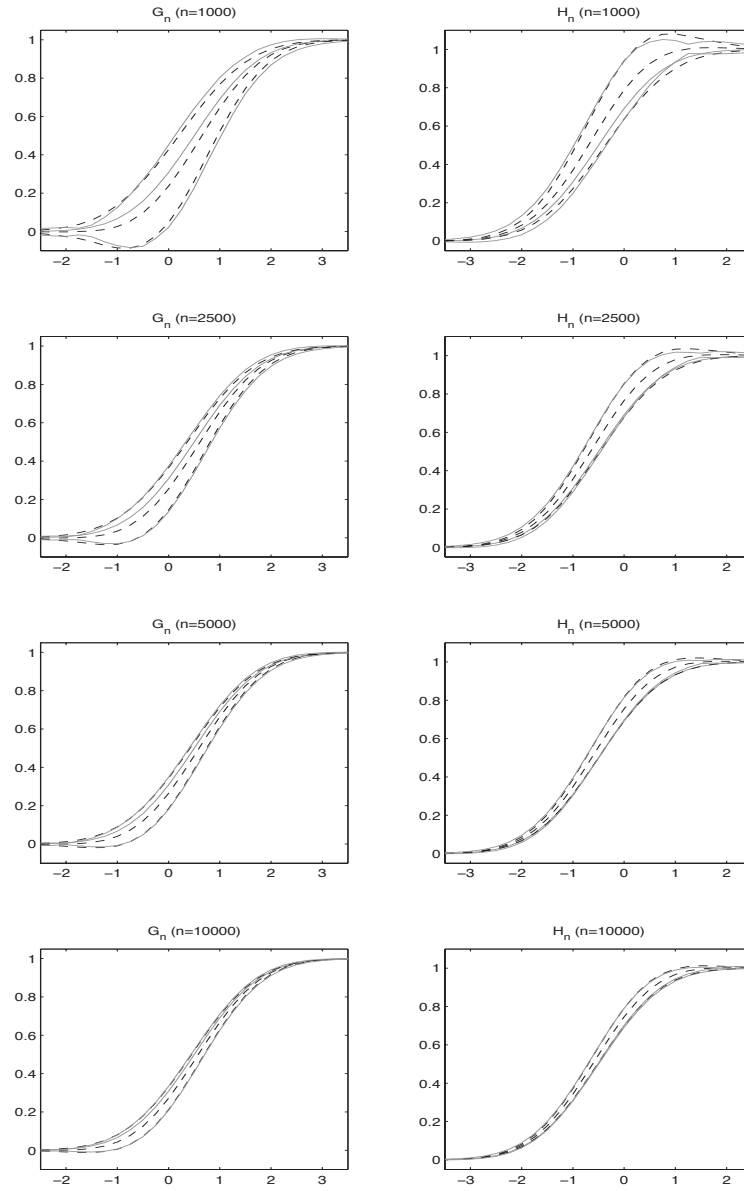
**Table 4.**  $\mu = .5$  and  $\beta = 0$ 

	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .75$										
500	.114	.886	.1131	−.1145	.1162	.1147	1.1992	1.2238	.9864	.9880
1,000	.090	.910	.0994	−.1017	.0842	.0855	1.1677	1.1599	.9416	.9406
2,500	.065	.935	.0840	−.0864	.0594	.0607	1.1175	1.0999	.8470	.8404
5,000	.051	.949	.0755	−.0758	.0454	.0463	1.1092	1.0920	.7448	.7342
10,000	.039	.961	.0671	−.0671	.0358	.0352	1.0815	1.0973	.6244	.6286
25,000	.028	.972	.0580	−.0573	.0257	.0259	1.0710	1.0601	.4374	.4584
$C = 1.25$										
500	.189	.811	.1451	−.1445	.0905	.0915	1.3579	1.3417	.9196	.9150
1,000	.149	.851	.1258	−.1271	.0673	.0682	1.2780	1.2632	.7896	.7808
2,500	.108	.892	.1075	−.1065	.0476	.0475	1.2051	1.2037	.5524	.5472
5,000	.084	.916	.0938	−.0947	.0369	.0371	1.1590	1.1555	.3806	.3672
10,000	.065	.935	.0832	−.0837	.0284	.0284	1.1426	1.1407	.2146	.2208
25,000	.047	.953	.0713	−.0714	.0205	.0208	1.1143	1.1002	.0740	.0740
$C = 1.75$										
500	.265	.735	.1704	−.1723	.0747	.0757	1.5656	1.5429	.8094	.7980
1,000	.209	.791	.1498	−.1500	.0573	.0580	1.4223	1.4072	.5592	.5578
2,500	.152	.849	.1255	−.1253	.0408	.0417	1.2970	1.2674	.2658	.2702
5,000	.118	.882	.1114	−.1095	.0323	.0318	1.2143	1.2267	.1076	.1160
10,000	.092	.908	.0968	−.0980	.0247	.0248	1.1889	1.1849	.0358	.0296
25,000	.065	.935	.0827	−.0827	.0178	.0180	1.1481	1.1324	.0034	.0030

**Figure 1.**  $\mu = 0$  and  $\beta = 5$  ( $C = .5$ )

**Figure 2.**  $\mu = 0$  and  $\beta = 2.5$  ( $C = .5$ )

**Figure 3.**  $\mu = 1$  and  $\beta = 0$  ( $C = .5$ )

**Figure 4.**  $\mu = .5$  and  $\beta = 0$  ( $C = .75$ )

## REFERENCES

- Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tabaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica* 80(5), 1977–2016.
- Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* 37, 3099–3132.
- Andrews, D. W. K. and M. M. A. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.
- Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review* 102, 94–130.
- Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature* 49, 3–71.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.
- Bollinger, C. R. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics* 73, 387–399.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014a). Estimating multivariate latent-structure models. Sciences Po Discussion Paper 2014-18.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014b). Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society - Series B*, forthcoming.
- Bordes, L., S. Mottelet, and P. Vandekerckhove (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics* 34, 1204–1232.
- Capitanio, A. (2010). On the approximation of the tail probability of the scalar skew-normal distribution. *METRON* 68, 299–308.
- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall, CRC Press.
- D’Haultfoeuille, X. and P. Février (2014). Identification of mixture models using support variations. Mimeo.
- D’Haultfoeuille, X. and A. Maurel (2013). Another look at identification at infinity of sample selection models. *Econometric Theory* 29, 213–224.
- Einmahl, J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *Journal of Statistical Planning and Inference* 32, 137–145.
- Einmahl, U. and D. Mason (1997). Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields* 107, 283–311.
- Frisch, R. (1934). Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics* 1, 255–294.
- Gassiat, E. and J. Rousseau (2014). Non parametric finite translation mixtures with dependent regime. *Bernoulli*, forthcoming.
- Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics Volume 14: Statistical Methods in Finance*. Elsevier.

- Hall, P. and X.-H. Zhou (2003). Nonparametric identification of component distributions in a multivariate mixture. *Annals of Statistics* 31, 201–224.
- Hamilton, J. D. (1989). A new approach to the analysis of nonstationary times series and the business cycle. *Econometrica* 57, 357–384.
- Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.
- Heckman, J. J. (1990). Varieties of selection bias. *American Economic Review* 80, 313–318.
- Henry, M., Y. Kitamura, and B. Salanié (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Paper 1767.
- Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5, 123–144.
- Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.
- Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77, 135–175.
- Khan, S. and E. Tamer (2010). Irregular identification, support conditions and inverse weight estimation. *Econometrica* 78, 2021–2042.
- Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.
- Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica* 74, 631–665.
- Schwarz, M. and S. Van Belleghem (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters* 80, 236–241.
- Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica* 68, 343–369.

## INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS\*

KOEN JOCHMANS<sup>1</sup> and MARC HENRY<sup>2</sup> and BERNARD SALANIÉ<sup>3</sup>

<sup>1</sup>*Sciences Po, 28 rue des Saint Pères, 75007 Paris, France.*

*E-mail:* [koen.jochmans@sciencespo.fr](mailto:koen.jochmans@sciencespo.fr)

<sup>2</sup>*The Pennsylvania State University, University Park, PA 16801, U.S.A.*

*E-mail:* [marc.henry@psu.edu](mailto:marc.henry@psu.edu)

<sup>3</sup>*Columbia University, 420 West 118th Street, New York, NY 10027, U.S.A.*

*E-mail:* [bsalanie@columbia.edu](mailto:bsalanie@columbia.edu)

Revised on March 27, 2015

Many econometric models can be analyzed as finite mixtures. We focus on two-component mixtures and we show that they are nonparametrically point identified by a combination of an exclusion restriction and tail restrictions. Our identification analysis suggests simple closed-form estimators of the component distributions and mixing proportions. We derive their asymptotic properties using results on tail empirical processes and we present a simulation study that documents their finite-sample performance.

*Keywords:* mixture model, nonparametric identification and estimation, tail empirical process.

## INTRODUCTION

The use of finite mixtures has a long history in applied econometrics. A non-exhaustive list of applications includes models with discrete unobserved heterogeneity, hidden Markov chains, and models with mismeasured discrete variables; see [Henry et al. \(2014\)](#) for a more extensive discussion of applications. Until recently, the literature on nonparametric identification of mixture models was sparse. Following the lead of [Hall and Zhou \(2003\)](#), several authors have analyzed multivariate mixtures; recent contributions are [Kasahara and Shimotsu \(2009\)](#), [Allman et al. \(2009\)](#), and [Bonhomme et al. \(2014a,b\)](#). There are fewer identifying restrictions available when the model of interest is univariate. [Bordes](#)

---

\*We are grateful to Arthur Lewbel and three referees for comments and suggestions, and to Victor Chernozhukov and Yuichi Kitamura for fruitful discussions. Parts of this paper were written while Henry was visiting the University of Tokyo Graduate School of Economics and while Salanié was visiting the Toulouse School of Economics. The hospitality of both institutions is gratefully acknowledged. Henry's research has received funding from SSHRC Grants 410-2010-242 and 435-2013-0292, and NSERC Grant 356491-2013. Jochmans' research has received funding from the SAB grant "Nonparametric estimation of finite mixtures". Some of the results presented here previously circulated as part of [Henry et al. \(2010\)](#), whose published version ([Henry et al. 2014](#)) only contains results on partial identification.



et al. (2006), for instance, provide such restrictions for location models with symmetric error distributions.

In this paper we give sufficient conditions that point-identify univariate component distributions and associated mixing proportions. The restrictions we rely on are most effective in two-component models. To simplify the analysis we focus on this case, like Hall and Zhou (2003) and Bordes et al. (2006). We comment briefly on mixtures with more components at the end of the paper. Our arguments are constructive, and we propose closed-form estimators for both the component distributions and the mixing proportions. We derive their large-sample properties and investigate their behavior in a simulation experiment.

The model we consider in this paper is characterized by an exclusion restriction and a tail-dominance assumption. Like Henry et al. (2014), we assume the existence of a source of variation that shifts the mixing proportions but leaves the component distributions unchanged. Such an assumption is natural in several important applications, such as measurement-error models (Mahajan 2006). In hidden Markov models, it follows directly from the model specification. The exclusion restriction is also implied by the conditional-independence restriction that underlies the results of Hall and Zhou (2003) and others on multivariate mixtures.

It was shown in Henry et al. (2014) that our exclusion restriction restricts both the component distributions and the mixing proportions to lie in a non-trivial set. Here we achieve point-identification by complementing the exclusion restriction with a restriction on the relative tail behavior of the component distributions. This restriction is quite natural in location models, for instance, but it can be motivated more generally. For example, regime-switching models typically feature regimes with different tail behavior. Alternatively, theoretical models can imply the required tail behavior; an example is the search and matching model of Shimer and Smith (2000), as explained in D'Haultfoeuille and Février (2014).

Our identification argument suggests plug-in estimators of the mixing proportions and the component distributions that are available in closed form. The estimators are based on ratios of intermediate quantiles, and their convergence rate is determined by the theory of tail empirical processes. As we rely on the tail behavior of the component

distributions to infer the mixing proportions, our estimators converge more slowly than the parametric rate. If the mixing proportions were known—or could be estimated at the parametric rate—the tail restrictions could be dispensed with and the implied estimator of the component distributions would also converge at the parametric rate.

Our estimators are consistent under very weak tail dominance assumptions. To control for asymptotic bias in their limit distribution we need to impose stronger requirements that prevent the tails of the components from vanishing too quickly. These assumptions rule out the Gaussian location model. Such thin-tailed distributions are known to be problematic for inference techniques that rely on tail behavior ([Khan and Tamer 2010](#)). However, we show that our assumptions apply to distributions with fatter tails, such as Pareto distributions.

Identification only requires that the variable subject to the exclusion restriction can take on two values. If it can take on more values, the model is overidentified and the specification can be tested. We derive the asymptotic properties of two statistics that can be used to respectively test the validity of the tail restrictions and that of the exclusion restriction.

The tail conditions we use to obtain nonparametric identification are related to the well-known identification-at-infinity argument of [Heckman \(1990\)](#); see also [D’Haultfoeulle and Maurel \(2013\)](#) for another approach. Other types of support restrictions have been used in related problems to establish identification. [Schwarz and Van Bellegem \(2010\)](#) imposed support restrictions in a semiparametric deconvolution problem to deal with measurement error in location models. [D’Haultfoeulle and Février \(2014\)](#) relied on a support condition as an alternative to completeness conditions ([Hu and Schennach 2008](#)) in multivariate mixture models.

The remainder of the paper is organized as follows. Section 1 describes the mixture model and proves identification. We rely on these results to construct estimators and derive their asymptotic properties in Section 2. Section 3 describes our specification tests. In Section 4 we conduct a Monte Carlo experiment that gives evidence on the small-sample performance of our methods. Finally, we conclude with some remarks on mixtures with more than two components.

## 1. MIXTURES WITH EXCLUSION AND TAIL RESTRICTIONS

Let  $(Y, X) \in \mathbb{R} \times \mathcal{X}$  be random variables. We assume throughout that our mixtures satisfy the following simple exclusion restriction.<sup>1</sup>

ASSUMPTION 1 (Mixture with exclusion).  *$F(y|x) \equiv \mathbb{P}(Y \leq y|X = x)$  decomposes as the two-component mixture*

$$F(y|x) = G(y) \lambda(x) + H(y) (1 - \lambda(x)) \quad (1.1)$$

*for distribution functions  $G : \mathbb{R} \mapsto [0, 1]$  and  $H : \mathbb{R} \mapsto [0, 1]$  and a function  $\lambda : \mathcal{X} \mapsto [0, 1]$  that maps values  $x$  into mixing proportions.*

The assumption that the component distributions do not depend on  $X$  embodies our exclusion restriction; see also Henry et al. (2014).

We complete the mixture model with the following assumption.

ASSUMPTION 2. *The mixing proportion  $\lambda$  is non-constant on  $\mathcal{X}$  and is bounded away from zero and one on  $\mathcal{X}$ .*

Non-constancy of  $\lambda$  gives the variable  $X$  relevance. Bounding  $\lambda$  away from zero and one implies that the mixture is irreducible.<sup>2</sup>

### 1.1. Motivating examples

Our first example has a long history in empirical work (Frisch 1934).

EXAMPLE 1 (Mismeasured treatments). *Let  $T$  denote a binary treatment indicator. Suppose that  $T$  is subject to classification error: rather than observing  $T$ , we observe misclassified treatment  $X$ . The distribution of the outcome variable  $Y$  given  $X = x$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1, X = x) \lambda(x) + \mathbb{P}(Y \leq y|T = 0, X = x) (1 - \lambda(x)),$$

---

<sup>1</sup>We omit conditioning variables throughout. The identification analysis extends straightforwardly. In principle, the distribution theory could be extended by using local empirical process results along the lines of Einmahl and Mason (1997). We postpone a detailed investigation into such an extension to future work.

<sup>2</sup>Note that irreducibility rules out the possibility of achieving identification of  $G$  and  $H$  via an identification-at-infinity argument, as in Heckman (1990) and Andrews and Schafgans (1998) for instance.

with  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . The usual ignorability assumption states that  $X$  and  $Y$  are independent given  $T$ . That is,

$$\mathbb{P}(Y \leq y|T = t, X = x) = \mathbb{P}(Y \leq y|T = t),$$

for  $t \in \{0, 1\}$ , in which case the decomposition of  $F(y|x)$  reduces to the model in (1.1) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$  and  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ . Also note that  $\lambda$  is non-constant unless misclassification in  $T$  is completely random.

The identification of treatment effects when the treatment indicator is mismeasured has received considerable attention, especially in the context of regression models (Bollinger 1996; Mahajan 2006; Lewbel 2007). Here, the conditional ignorability assumption that validates our exclusion restriction relies on non-differential misclassification error. It has been routinely used elsewhere (Carroll et al. 2006).

Our second example deals with regime-switching models, also referred to as hidden Markov models. These models cover switching regressions, which have been used in a variety of settings (see, e.g., Heckman 1974, Hamilton 1989), as well as several versions of stochastic-volatility models (Ghysels et al. 1996).

EXAMPLE 2 (Hidden Markov model). *Let  $\mathbf{Y} = (Y_1, \dots, Y_T)'$  be a time series of outcome variables. A hidden Markov model for the dependency structure in these data assumes that there is a discrete latent series of state variables  $\mathbf{S} = (S_1, \dots, S_T)'$  having Markov dependence, that the variables in  $\mathbf{Y}$  are jointly independent given  $\mathbf{S}$ , and that*

$$\mathbb{P}(Y_t \leq y_t|\mathbf{S} = \mathbf{s}) = \mathbb{P}(Y_t \leq y_t|S_t = s_t).$$

*To see that such a model fits (1.1), assume that there are two latent states 0 and 1 and (for notational simplicity) that  $\mathbf{S}$  has first-order Markov dependence. Denote  $\mathbf{X} = (Y_1, \dots, Y_{t-1})'$ . Then*

$$F(y_t|\mathbf{x}) = \mathbb{P}(Y_t \leq y_t|S_t = 1)\mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x}) + \mathbb{P}(Y_t \leq y_t|S_t = 0)\mathbb{P}(S_t = 0|\mathbf{X} = \mathbf{x}),$$

*which fits our setup. Moreover,  $\lambda(\mathbf{x}) = \mathbb{P}(S_t = 1|\mathbf{X} = \mathbf{x})$  does vary with  $\mathbf{x}$ , unless the outcomes are independent of the latent states.*

In this example, the exclusion restriction follows directly from the Markovian structure of the regime-switching model. [Gassiat and Rousseau \(2014\)](#) obtained nonparametric identification in location models when the matrix of transition probabilities of the Markov chain has full rank. The approach presented here delivers nonparametric identification in a much broader range of models.

Our third example links [\(1.1\)](#) to the recent literature on multivariate mixtures building on [Hall and Zhou \(2003\)](#).

**EXAMPLE 3 (Multivariate mixture).** *Suppose  $Y$  and  $X$  are two measurements that are independent conditional on a latent binary factor  $T$ :*

$$\begin{aligned}\mathbb{P}(Y \leq y, X \leq x) &= \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(X \leq x|T = 1) \mathbb{P}(T = 1) \\ &\quad + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(X \leq x|T = 0) \mathbb{P}(T = 0).\end{aligned}$$

*Then the conditional distribution of the  $Y$  given  $X$  is*

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1) \mathbb{P}(T = 1|X = x) + \mathbb{P}(Y \leq y|T = 0) \mathbb{P}(T = 0|X = x).$$

*This is of the form in [\(1.1\)](#) with  $G(y) = \mathbb{P}(Y \leq y|T = 1)$ ,  $H(y) = \mathbb{P}(Y \leq y|T = 0)$ , and  $\lambda(x) = \mathbb{P}(T = 1|X = x)$ . Note that the bivariate mixture model implies that the distribution of  $X$  given  $Y$  decomposes in the same way.*

[Hall and Zhou \(2003\)](#) showed that multivariate two-component mixtures with conditional independence restrictions are nonparametrically identified from data on three or more measurements and are set identified from data on only two measurements. The results we derive below imply that two measurements can also yield point identification under tail restrictions.

## 1.2. Identification

We show below that both the mixture components  $G, H$  and the mixing proportions  $\lambda$  are identified under the following dominance condition on the tails of the component distributions.

ASSUMPTION 3 (Tail dominance).

(i) The left tail of  $G$  is thinner than the left tail of  $H$ , i.e.,

$$\lim_{y \downarrow -\infty} \frac{G(y)}{H(y)} = 0.$$

(ii) The right tail of  $G$  is thicker than the right tail of  $H$ , i.e.,

$$\lim_{y \uparrow +\infty} \frac{1 - H(y)}{1 - G(y)} = 0.$$

Tail dominance is natural in location models.

EXAMPLE 4 (Location models). Suppose that  $Y = \mu(T) + U$ , where  $T$  is a binary indicator and  $U \sim F$ , independent of  $T$ . Then (1.1) yields

$$F(y|x) = F(y - \mu(1)) \mathbb{P}(T = 1|X = x) + F(y - \mu(0)) \mathbb{P}(T = 0|X = x).$$

Suppose that  $\mu(0) < \mu(1)$ , that  $F$  is absolutely continuous, and that the derivatives of  $-\ln(1 - F(u))$  and  $-\ln F(u)$  grow without bound as  $u \uparrow +\infty$  and  $u \downarrow -\infty$ , respectively. Then Assumption 3 holds with  $G(y) = F(y - \mu(1))$  and  $H(y) = F(y - \mu(0))$ .

**Proof.** We prove Assumption 3(ii). Let  $\varphi(u) \equiv -\ln(1 - F(u))$  and denote  $\varphi'$  its first derivative. Then

$$\frac{1 - F(y - \mu(0))}{1 - F(y - \mu(1))} = \exp(\varphi(y - \mu(1)) - \varphi(y - \mu(0))) = \exp(-\varphi'(y^*)(\mu(1) - \mu(0))),$$

where  $y^*$  lies between  $y - \mu(1)$  and  $y - \mu(0)$ . Since  $\varphi'$  increases without bound as  $y \uparrow +\infty$  and  $\mu(1) > \mu(0)$ , the expression on the right-hand side tends to zero as  $y$  increases. Assumption 3(i) can be verified in the same way.  $\square$

We now show that, combined, our exclusion restriction and tail dominance identify all elements of the mixture model.

THEOREM 1 (Identification). Under Assumptions 1–3,  $G$ ,  $H$ , and  $\lambda$  are identified.

**Proof.** The proof is constructive. Fix  $x' \in \mathcal{X}$  and choose  $x'' \in \mathcal{X}$  so that  $\lambda(x') \neq \lambda(x'')$ . Then re-arranging (1.1) gives

$$\begin{aligned} \frac{F(y|x')}{F(y|x'')} &= \frac{1 + \lambda(x') (G(y)/H(y) - 1)}{1 + \lambda(x'') (G(y)/H(y) - 1)}, \\ \frac{1 - F(y|x')}{1 - F(y|x'')} &= \frac{\lambda(x') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x'))}{\lambda(x'') + ((1 - H(y))/(1 - G(y))) (1 - \lambda(x''))}. \end{aligned}$$

Taking limits, Assumption 3 further implies that

$$\begin{aligned} \zeta^-(x', x'') &\equiv \lim_{y \downarrow -\infty} \frac{F(y|x')}{F(y|x'')} = \frac{1 - \lambda(x')}{1 - \lambda(x'')}, \\ \zeta^+(x', x'') &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|x')}{1 - F(y|x'')} = \frac{\lambda(x')}{\lambda(x'')}. \end{aligned} \tag{1.2}$$

These two equations can be solved for the mixing proportion at  $x'$ , yielding

$$\lambda(x') = \frac{1 - \zeta^-(x'', x')}{\zeta^+(x'', x') - \zeta^-(x'', x')}. \tag{1.3}$$

Since  $\lambda$  is non-constant, for any  $x' \in \mathcal{X}$  there exists a  $x'' \in \mathcal{X}$  for which such a system of equations can be constructed. The function  $\lambda$  is therefore identified on its entire support.

To establish identification of  $G$  and  $H$ , first note that

$$G(y) - H(y) = \frac{F(y|x'') - F(y|x')}{\lambda(x'') - \lambda(x')} \tag{1.4}$$

follows from (1.1). Then, evaluating (1.1) in  $x''$  and re-arranging the resulting expression for  $F(y|x'')$  gives

$$H(y) = F(y|x'') - (G(y) - H(y)) \lambda(x'') = F(y|x'') - \frac{\lambda(x'')}{\lambda(x'') - \lambda(x')} (F(y|x'') - F(y|x')),$$

which is identified. Furthermore, using (1.2) we can write

$$H(y) = F(y|x'') - \frac{1}{1 - \zeta^+(x', x'')} (F(y|x'') - F(y|x')). \tag{1.5}$$

Plugging this expression for  $H(y)$  back into the mixture representation of  $F(y|x'')$  as in (1.1) further yields

$$G(y) = F(y|x'') - \frac{1}{1 - \zeta^-(x', x'')} (F(y|x'') - F(y|x')), \tag{1.6}$$

again using (1.2). This shows that both component distributions are identified, concluding the proof.  $\square$

If we only assume one-sided tail dominance, then either  $G$  or  $H$  remains identified.

**COROLLARY 1** (One-sided tail dominance). *Under Assumptions 1 and 2,  $G$  is identified if Assumption 3(i) holds and  $H$  is identified if Assumption 3(ii) holds.*

**Proof.** We consider identification of  $H$ . Let  $x', x''$  be as in the proof of Theorem 1. Under Assumption 3(ii) we can still determine  $\zeta^+(x', x'') = \lambda(x')/\lambda(x'')$ , from which we can learn the ratio  $1/(1 - \zeta^+(x'', x'))$ . Together with (1.5) this yields  $H$ . This concludes the proof of the corollary.  $\square$

The following example illustrates the usefulness of Corollary 1.

**EXAMPLE 5** (Stochastic volatility). *Consider a two-regime stochastic volatility model, which is a special case of Example 2. Assume that the outcome variable  $Y$  has mean zero and conditional variance*

$$T\sigma_G^2 + (1 - T)\sigma_H^2$$

*for positive constants  $\sigma_G^2$  and  $\sigma_H^2$ . Suppose that  $\sigma_G^2 > \sigma_H^2$ . Then  $G$  is the distribution associated with a regime that is characterized by relatively higher volatility. In this case, both tails of  $G$  dominate those of  $H$ . Hence, in Assumption 3, Condition (ii) holds but Condition (i) fails. Nevertheless, the distribution of the lower-volatility regime, namely  $H$ , remains identified.*

Our identification result suggests plug-in estimators of the mixing proportions and the component distributions. The proof of Theorem 1, and Equations (1.3) and (1.5)–(1.6) in particular, further show that our mixture model yields overidentifying restrictions as soon as the instrument can take on more than two values. We turn to estimation in the next section. Afterwards we discuss the construction of specification tests under overidentification.



## 2. ESTIMATION

To motivate the construction of our estimators, we first note that the structure of the model in (1.1) continues to hold when we aggregate across  $x$ . Extending our notation to

$$F(y|A) \equiv \mathbb{P}(Y \leq y|X \in A), \quad \lambda(A) \equiv \sum_{x \in A} \lambda(x) \mathbb{P}(X = x|X \in A),$$

for any  $A \subset \mathcal{X}$ , we have

$$F(y|A) = G(y) \lambda(A) + H(y) (1 - \lambda(A)), \quad (2.1)$$

which is of the same form as (1.1). Furthermore, the proof of Theorem 1 continues to go through for (2.1); replacing  $x'$  with  $A$  and  $x''$  with  $\mathcal{X} - A$  does not alter the argument.

We will assume from now on that  $X$  is discrete. As will become apparent, this only entails a loss of generality for the estimation of the function  $\lambda$ , as our estimator will only yield a discretized approximation to it. Extending our results to continuous  $X$  would complicate the exposition greatly and we feel that it would only distract from our main argument.

We will work under the following sampling condition.

ASSUMPTION 4.  $(Y_1, X_1), \dots, (Y_n, X_n)$  is a random sample on  $(Y, X)$ .

For each  $A \subset \mathcal{X}$ , let

$$F_n(y|A) \equiv n_A^{-1} \sum_{i=1}^n 1\{Y_i \leq y, X_i \in A\},$$

where  $n_A \equiv \sum_{i=1}^n 1\{X_i \in A\}$ .

For each pair of disjoint subsets  $A, B$  of  $\mathcal{X}$  we can generalize (1.2) to

$$\begin{aligned} \zeta^-(A, B) &\equiv \lim_{y \downarrow -\infty} \frac{F(y|A)}{F(y|B)} = \frac{1 - \lambda(A)}{1 - \lambda(B)}, \\ \zeta^+(A, B) &\equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|A)}{1 - F(y|B)} = \frac{\lambda(A)}{\lambda(B)}. \end{aligned} \quad (2.2)$$

For any subsample of size  $m$  and integers  $\iota_m$  and  $\kappa_m$ , let  $\ell_m$  and  $r_m$  denote the  $(\iota_m + 1)$ th and  $(m - \kappa_m)$ th order statistics of  $Y$  in this subsample. We estimate the quantities in (2.2) by

$$\zeta_n^-(A, B) \equiv \frac{F_n(\ell_{n_B}|A)}{F_n(\ell_{n_B}|B)}, \quad \zeta_n^+(A, B) \equiv \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)}, \quad (2.3)$$

respectively. In our asymptotic theory, we will choose  $\iota_{n_B}$  and  $\kappa_{n_B}$  so that  $\ell_{n_B} \downarrow -\infty$  and  $r_{n_B} \uparrow +\infty$  as  $n \uparrow +\infty$  at an appropriate rate.

Estimators of both the mixing proportions and the component distributions follow readily along the lines of the proof of Theorem 1; see below. Since their asymptotic distribution will be driven by the large-sample behavior of the estimators of the quantities in (2.3), we start by deriving the statistical properties of these estimators.

### 2.1. Asymptotic theory for intermediate quantiles

Throughout this section we fix disjoint sets  $A, B$  and consider the asymptotic behavior of the estimators in (2.3).

Consistency only requires the following rate conditions.

ASSUMPTION 5 (Order statistics).  $\iota_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  and  $\kappa_{n_B}/\sqrt{n_B \ln \ln n_B} \uparrow +\infty$  as  $n \uparrow +\infty$ .

THEOREM 2 (Consistency). *If Assumptions 1–5 hold,*

$$\zeta_n^-(A, B) \xrightarrow{P} \zeta^-(A, B), \quad \zeta_n^+(A, B) \xrightarrow{P} \zeta^+(A, B),$$

as  $n \uparrow +\infty$ .

**Proof.** We prove the theorem for  $\zeta_n^+$ ; the proof for  $\zeta_n^-$  follows in a similar fashion. Write

$$\zeta_n^+ - \zeta^+ = (\zeta_n^+ - \zeta^{\kappa_{n_B}}) + (\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.4)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . For the second right-hand side term in (2.4) we have

$$\zeta^{\kappa_{n_B}} - \zeta^+ = \left( \frac{\lambda(A) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(A))}{\lambda(B) + \frac{1-H(r_{n_B})}{1-G(r_{n_B})}(1-\lambda(B))} - \frac{\lambda(A)}{\lambda(B)} \right) = O_p\left(\frac{1-H(r_{n_B})}{1-G(r_{n_B})}\right) = o_p(1),$$

by Assumptions 3(ii) and 5. To deal with the first right-hand side term in (2.4), recall that

$$\zeta_n^+ - \zeta^{\kappa_{n_B}} = \frac{1 - F_n(r_{n_B}|A)}{1 - F_n(r_{n_B}|B)} - \frac{1 - F(r_{n_B}|A)}{1 - F(r_{n_B}|B)}.$$

Letting  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$  we thus have that

$$\begin{aligned} \zeta_n^+ - \zeta^{\kappa_{n_B}} &= \frac{(1 - F(r_{n_B}|A))\mathbb{G}_n(r_{n_B}|B)/\sqrt{n_B} - (1 - F(r_{n_B}|B))\mathbb{G}_n(r_{n_B}|A)/\sqrt{n_A}}{(1 - F_n(r_{n_B}|B))(1 - F(r_{n_B}|B))} \\ &= \frac{\sqrt{n_B}}{\kappa_{n_B}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right) \\ &= O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right), \end{aligned}$$

where the second equality uses  $1 - F_n(r_{n_B}|B) = \kappa_{n_B}/n_B$  and the last one follows by the law of the iterated logarithm for empirical processes. Thus, from Assumption 5 it follows that  $|\zeta_n^+ - \zeta^{\kappa_{n_B}}| = o_p(1)$ . This completes the proof.  $\square$

Deriving the limit distribution requires some more care, and three more assumptions. We first impose the following regularity condition on the component distributions.

ASSUMPTION 6. *G and H are absolutely continuous on  $\mathbb{R}$ .*

This assumption is very weak. Note that we do not require the existence of any moments of the component distributions. Thus, our results apply to heavy-tailed distributions such as Cauchy and Pareto distributions.

We will also complement Assumption 5 with an additional rate condition.

ASSUMPTION 7 (Order statistics (cont'd.)).  *$\iota_{n_B}/n_B \downarrow 0$  and  $\kappa_{n_B}/n_B \downarrow 0$  as  $n \uparrow +\infty$ .*

Where Assumption 5 required the order statistics to grow to ensure consistency, this assumption bounds this growth rate so that appropriately scaled versions of  $\zeta_n^+$  and  $\zeta_n^-$  have a limit distribution.

Finally, we will use an additional condition on the relative tails of the component distributions.

ASSUMPTION 8 (Tail rates).

- (i)  $G(\ell_{n_B})/H(\ell_{n_B}) = o_p(1/\sqrt{\iota_{n_B}})$ ; and
- (ii)  $(1 - H(r_{n_B}))/ (1 - G(r_{n_B})) = o_p(1/\sqrt{\kappa_{n_B}})$ .

Assumption 8 rules out distributions whose tails vanish too quickly and ensures that the limit distributions are free of asymptotic bias. We comment on Assumption 8 after we derive the limit distributions of our estimators.

Let  $\rho_{A,B} \equiv \mathbb{P}(X \in B)/\mathbb{P}(X \in A)$ . Note that  $0 < \rho_{A,B} < +\infty$  because of random sampling. Introduce

$$\begin{aligned}\sigma_-^2(A, B) &\equiv \zeta^-(A, B)^2 + \rho_{A,B} \zeta^-(A, B), \\ \sigma_+^2(A, B) &\equiv \zeta^+(A, B)^2 + \rho_{A,B} \zeta^+(A, B),\end{aligned}$$

Theorem 2 provides the asymptotic properties of the estimators in (2.2) and is the main building block for our subsequent results.

**THEOREM 3** (Asymptotic normality). *If Assumptions 1–8 hold, then as  $n \uparrow +\infty$ ,*

$$\begin{aligned}\sqrt{\iota_{n_B}}(\zeta_n^-(A, B) - \zeta^-(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, B)), \\ \sqrt{\kappa_{n_B}}(\zeta_n^+(A, B) - \zeta^+(A, B)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, B));\end{aligned}$$

*and these two estimators are asymptotically independent.*

**Proof.** We focus on the limit behavior of  $\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+)$  here; the proof of the result for  $\sqrt{\iota_n}(\zeta_n^- - \zeta^-)$  follows similar lines.

As in the proof of Theorem 2, write

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^+) = \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) + \sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+), \quad (2.5)$$

for  $\zeta^{\kappa_{n_B}} \equiv (1 - F(r_{n_B}|A))/(1 - F(r_{n_B}|B))$ . Assumption 8 implies that

$$\sqrt{\kappa_{n_B}}(\zeta^{\kappa_{n_B}} - \zeta^+) = \sqrt{\kappa_{n_B}} O_p\left(\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})}\right) = o_p(1).$$

Hence, the second right-hand side term in (2.5) is asymptotically negligible.

We now turn to the first term in (2.5). From the proof of Theorem 2 we have that

$$\sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) = \sqrt{\frac{n_B}{\kappa_{n_B}}} \left( \zeta^{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) - \sqrt{\frac{n_B}{n_A}} \mathbb{G}_n(r_{n_B}|A) \right),$$

where  $\mathbb{G}_n(y|S) \equiv \sqrt{n_S}(F_n(y|S) - F(y|S))$  for any  $S \subset \mathcal{X}$ . Let  $\alpha_n(u) \equiv \sqrt{n}(\mathcal{U}_n(u) - u)$  for  $\mathcal{U}_n$  the empirical cumulative distribution of an i.i.d. sample of size  $n$  from a uniform distribution on  $[0, 1]$ . By Assumption 6,  $F(y|S)$  is continuous in  $y$  for all  $S \subset \mathcal{X}$ . Therefore

$$\mathbb{G}_n(y|A) = \alpha_{n_A}(1 - F(y|A)) \text{ and } \mathbb{G}_n(y|B) = \alpha_{n_B}(1 - F(y|B))$$

by an application of the probability integral transform. Hence, we may write

$$\begin{aligned} \sqrt{\kappa_{n_B}}(\zeta_n^+ - \zeta^{\kappa_{n_B}}) &= \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \\ &\quad - \sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A}(1 - F(r_{n_B}|A)). \end{aligned} \quad (2.6)$$

We study the asymptotic behavior of each of the right-hand side terms in turn.

Start with the first right-hand side term in (2.6). From the definition of the order statistic  $r_{n_B}$ , we find by adding and subtracting  $F_n(r_{n_B}|B)$  that

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} \left( 1 + \frac{\sqrt{n_B}}{\kappa_{n_B}} \mathbb{G}_n(r_{n_B}|B) \right);$$

or, defining  $\varepsilon_n \equiv -\sqrt{n_B}/\kappa_{n_B} \mathbb{G}_n(r_{n_B}|B)$ ,

$$1 - F(r_{n_B}|B) = \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Therefore we can write

$$\zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{\kappa_{n_B}}} \alpha_{n_B}(1 - F(r_{n_B}|B)) = \sqrt{2} \zeta^{\kappa_{n_B}} \sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right). \quad (2.7)$$

By the law of the iterated logarithm together with Assumption 5,

$$\varepsilon_n = -\frac{\sqrt{n_B}}{\kappa_{n_B}} O_{a.s.} \left( \sqrt{\ln \ln n_B} \right) = O_{a.s.} \left( \frac{\sqrt{n_B \ln \ln n_B}}{\kappa_{n_B}} \right) = o_{a.s.}(1).$$

Hence,  $(1 - \varepsilon_n)/2$  converges almost surely to  $1/2$ , and  $(1 - \varepsilon_n)/2 \in (0, 1)$  for  $n$  large enough. We may then apply Theorem 2.1 in Einmahl (1992) to establish the convergence in distribution of  $\sqrt{\frac{n_B}{2\kappa_{n_B}}} \alpha_{n_B} \left( \frac{2\kappa_{n_B}}{n_B} \frac{1 - \varepsilon_n}{2} \right)$  to a normal random variable with mean zero and variance  $1/2$ . This, together with Equation (2.7) and an application of Slutsky's theorem, then implies that

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \zeta^{\kappa_{n_B}} \alpha_{n_B}(1 - F(r_{n_B}|B)) \xrightarrow{d} \zeta^+ Z_B^+, \quad (2.8)$$

where  $Z_B^+$  is a standard normal random variable.

Now turn to the second right-hand side term in (2.6). First observe that

$$1 - F(r_{n_B}|A) = \zeta^{\kappa_{n_B}} (1 - F(r_{n_B}|B)) = \zeta^{\kappa_{n_B}} \frac{\kappa_{n_B}}{n_B} (1 - \varepsilon_n).$$

Using  $\rho_{A,B} = \lim_{n \uparrow +\infty} n_B/n_A$ , this gives

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) = \sqrt{2\rho_{A,B} \zeta^+} \sqrt{\frac{n_A}{2\tilde{\kappa}_{n_A}}} \alpha_{n_A} \left( \frac{2\tilde{\kappa}_{n_A}}{n_A} \frac{1 - \varepsilon_n}{2} \right) + o_p(1),$$

where  $\tilde{\kappa}_{n_A} \equiv (\kappa_{n_B} \zeta^{\kappa_{n_B}})/(n_B/n_A)$ . As  $\tilde{\kappa}_{n_A}$  satisfies Assumption (1.5) of Theorem 2.1 in Einmahl (1992) we may apply his theorem again to obtain

$$\sqrt{\frac{n_B}{\kappa_{n_B}}} \sqrt{\frac{n_B}{n_A}} \alpha_{n_A} (1 - F(r_{n_B}|A)) \xrightarrow{d} \sqrt{\rho_{A,B} \zeta^+} Z_A^+, \quad (2.9)$$

where  $Z_A^+$  is a standard-normal random variable which, by random sampling, is independent of  $Z_B^+$ .

Combining (2.6) with (2.8) and (2.9) then gives

$$\sqrt{\kappa_{n_B}} (\zeta_n^+ - \zeta^{\kappa_{n_B}}) \xrightarrow{d} \zeta^+ Z_B^+ - \sqrt{\rho_{A,B} \zeta^+} Z_A^+,$$

as claimed. This concludes the proof.  $\square$

We finish this section with two examples that specialize Assumption 8 to densities with log-concave tails and Pareto tails, respectively. In both cases, Assumption 8 is implied by the rate conditions in Assumption 7.

EXAMPLE 6 (Log-concave tails). *Suppose that  $G$  and  $H$  have log-concave tails; and for notational simplicity, assume that*

$$-\ln(1 - G(y)) \sim \left( \frac{y}{\sigma_G^+} \right)^{\alpha_G^+}, \quad -\ln(1 - H(y)) \sim \left( \frac{y}{\sigma_H^+} \right)^{\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for real numbers  $\alpha_G^+, \alpha_H^+ > 1$  and  $\sigma_G^+, \sigma_H^+ > 0$ , and

$$-\ln G(y) \sim \left( \frac{-y}{\sigma_G^-} \right)^{\alpha_G^-}, \quad -\ln H(y) \sim \left( \frac{-y}{\sigma_H^-} \right)^{\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

for real numbers  $\alpha_G^-, \alpha_H^- > 1$  and  $\sigma_G^-, \sigma_H^- > 0$ . Then Assumption 7 implies Assumption 8 if both

(i)  $\alpha_G^+ < \alpha_H^+$ , or  $\alpha_G^+ = \alpha_H^+$  and  $\sigma_G^+ > \sigma_H^+$ ; and

(ii)  $\alpha_G^- > \alpha_H^-$ , or  $\alpha_G^- = \alpha_H^-$  and  $\sigma_G^- < \sigma_H^-$

hold.

**Proof.** We verify the second rate; the first follows similarly. Throughout, fix the set  $B$ . Assumptions 3(ii) and 7 imply that

$$\begin{aligned} 1 - F(r_{n_B}|B) &= (1 - G(r_{n_B})) \lambda(B) + (1 - H(r_{n_B})) (1 - \lambda(B)) \\ &= (1 - G(r_{n_B})) (\lambda(B) + o_p(1)). \end{aligned}$$

Further, because  $\kappa_{n_B}/n_B = 1 - F_n(r_{n_B}|B)$ , adding and subtracting  $F(r_{n_B}|B)$  gives

$$\begin{aligned} \frac{\kappa_{n_B}}{n_B} &= (1 - F(r_{n_B}|B)) + (F_n(r_{n_B}|B) - F(r_{n_B}|B)) \\ &= (1 - F(r_{n_B}|B)) + O_{a.s.}(\sqrt{(\ln \ln n_B)/n_B}). \end{aligned}$$

Because  $(\ln \ln n_B)/n_B \rightarrow 0$ , put together, we find

$$\frac{\kappa_{n_B}}{n_B} = C (1 - G(r_{n_B})) (1 + o_p(1))$$

for some constant  $C$ . With  $G$  and  $H$  having log-concave tails, it then follows from this expression that, asymptotically,  $r_{n_B}$  behaves like  $\alpha_G^+ \sqrt[n_B]{\ln n_B}$ . So, because

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} \sim \exp \left\{ \left( \frac{r_{n_B}}{\sigma_G^+} \right)^{\alpha_G^+} - \left( \frac{r_{n_B}}{\sigma_H^+} \right)^{\alpha_H^+} \right\},$$

we have that

$$\frac{1 - H(r_{n_B})}{1 - G(r_{n_B})} = \begin{cases} O_p \left( \exp(-(\ln n_B)^{\alpha_H^+/\alpha_G^+}) \right) & \text{if } \alpha_H^+ > \alpha_G^+ \\ O_p(1/n_B) & \text{if } \alpha_H^+ = \alpha_G^+ \text{ and } \sigma_H^+ < \sigma_G^+ \end{cases},$$

from which the conclusion follows.  $\square$

Example 6 does not cover location models with log-concave distributions for which the  $\alpha$  and  $\sigma$  parameters of  $H$  equal those of  $G$ . This includes the location model with Gaussian errors, for which  $\alpha = 2$  and  $\sigma$  is the common standard error. While our estimator remains

consistent in such cases, we do not know of general results on tail empirical processes that would yield the asymptotic distribution of the estimator in this knife-edge case. To assess the extent to which the failure of Assumption 8 may play a role for inference, our simulation experiments in Section 4 include a Gaussian location model.

EXAMPLE 7 (Pareto tails). *Let  $C$  denote a generic constant. Suppose that  $G$  and  $H$  have Pareto tails, i.e.,*

$$(1 - G(y)) \sim C y^{-\alpha_G^+}, \quad (1 - H(y)) \sim C y^{-\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for positive real numbers  $\alpha_H^+ > \alpha_G^+$  and

$$G(y) \sim C (-y)^{-\alpha_G^-}, \quad H(y) \sim C (-y)^{-\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

for positive real numbers  $\alpha_G^- < \alpha_H^-$ . Then Assumption 7 implies Assumption 8.

**Proof.** The argument is very similar to the one that was used to verify Example 6. We treat the right tail only. We have

$$\frac{\kappa_{n_B}}{n_B} = (1 - G(r_{n_B})) (1 + o_p(1)) = C r^{-\alpha_G^+} (1 + o_p(1)).$$

Assumption 8 requires that  $(1 - H(r_{n_B})) / (1 - G(r_{n_B})) = o(1/\sqrt{\kappa_n})$ , that is, that  $r_{n_B}^{\alpha_G^+ - \alpha_H^+} = o_p(1/\sqrt{\kappa_{n_B}})$ . This rate condition is satisfied when

$$\left( \frac{n_B}{\kappa_{n_B}} \right)^{\frac{\alpha_G^+ - \alpha_H^+}{\alpha_G^+}} = o_p \left( \frac{1}{\sqrt{\kappa_{n_B}}} \right),$$

which can be achieved by setting  $\kappa_{n_B} = o(n_B^{\gamma^+})$  for

$$\gamma^+ \equiv \frac{\alpha_H^+ - \alpha_G^+}{\alpha_H^+ - \alpha_G^+/2}. \quad (2.10)$$

This condition is weaker than Assumption 7 and is therefore implied by it.  $\square$

Example 7 shows that our methods are well suited to deal with Pareto tails. Pareto tails show up in many economic applications. A time-honored example is income and wealth distributions (Atkinson et al. 2011), which are often modeled as a log-normal for



most quantiles, combined with a Pareto right tail. More generally, “power laws” have become a popular tool in finance, in studies of firm growth, and in urban economics (see [Gabaix 2009](#) for a recent survey, and [Acemoglu et al. 2012](#) for an application to business cycles.) Many recent models of monopolistic competition, as used in international trade for instance, also assume that productivities are Pareto-distributed ([Arkolakis et al. 2012](#)).

Identification only requires that the tail index of  $H$  be larger than that of  $G$ , that is,  $\alpha_H > \alpha_G$ . Denote  $c = \alpha_H/\alpha_G > 1$ . Equation (2.10) then gives a convergence rate arbitrarily close to  $n^{-\beta/2}$  for  $\beta = 2(c-1)/(2c-1)$ . For example, if  $c = 2$  then  $\beta = 2/3$  and our estimators will converge slightly more slowly than  $n^{-1/3}$ . However, as  $c$  increases,  $\beta$  rapidly converges to one, and our estimators will converge at close to the  $n^{-1/2}$  parametric rate.

## 2.2. Mixing proportions

Fix  $x \in \mathcal{X}$  and consider estimating  $\lambda(x)$ . Set  $A = \mathcal{X} - x$  and  $B = x$  in (2.2) and solve for  $\lambda(x)$  to get

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}.$$

The mixing proportion  $\lambda$  need not be a strictly monotonic function. Estimating  $\lambda(x)$  by an average of plug-in estimates of (1.3) could therefore be problematic, as the denominator in (1.3) can be zero or be arbitrarily close to it for some pairs of values  $(x', x'')$ .

We instead estimate the mixing proportion at  $X = x$  by a plug-in estimator based on (2.3), that is,

$$\lambda_n(x) \equiv \frac{1 - \zeta_n^-(A, x)}{\zeta_n^+(A, x) - \zeta_n^-(A, x)}.$$

This estimator uses observations with  $X_i \neq x$  in a way that immunizes it against small or zero denominators.

To present the asymptotic variance of this estimator we need to define

$$\begin{aligned} d^-(x) &\equiv \frac{1 - \zeta^+(A, x)}{(\zeta^+(A, x) - \zeta^-(A, x))^2}, \\ d^+(x) &\equiv \frac{\zeta^-(A, x) - 1}{(\zeta^+(A, x) - \zeta^-(A, x))^2}. \end{aligned} \tag{2.11}$$

The speed of convergence and the asymptotic distribution of the  $\lambda_n(x)$  depend on the ratio  $c_x \equiv \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

**THEOREM 4** (Mixing proportions). *Under the conditions of Theorem 2,*

$$|\lambda_n(x) - \lambda(x)| = o_p(1)$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, d^-(x)^2 \sigma_-^2(A, x) + c_x d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x < +\infty, \\ \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &\xrightarrow{d} \mathcal{N}(0, c_x^{-1} d^-(x)^2 \sigma_-^2(A, x) + d^+(x)^2 \sigma_+^2(A, x)) \quad \text{if } c_x > 0, \end{aligned}$$

as  $n \uparrow +\infty$ .

**Proof.** The consistency claim follows directly from Theorem 2 by an application of the continuous mapping theorem.

To establish the asymptotic distribution, note that Theorem 2 states that

$$\begin{aligned} \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_-^2(A, x)), \\ \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) &\xrightarrow{d} \mathcal{N}(0, \sigma_+^2(A, x)), \end{aligned}$$

and that  $\zeta_n^-(x)$  and  $\zeta_n^+(x)$  are asymptotically independent. An expansion around  $\zeta^-(A, x)$  and  $\zeta^+(A, x)$  then yields

$$\begin{aligned} \sqrt{\iota_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(A, x) - \zeta^-(A, x)) \\ &\quad + d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(A, x) - \zeta^+(A, x)) \sqrt{\frac{\iota_{n_x}}{\kappa_{n_x}}} + o_p(1), \end{aligned}$$

which has the limit distribution stated in the theorem if  $c_x$  is finite. Also, by the same argument,

$$\begin{aligned} \sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda(x)) &= d^+(x) \sqrt{\kappa_{n_x}}(\zeta_n^+(x) - \zeta^+(x)) \\ &\quad + d^-(x) \sqrt{\iota_{n_x}}(\zeta_n^-(x) - \zeta^-(x)) \sqrt{\frac{\kappa_{n_x}}{\iota_{n_x}}} + o_p(1) \end{aligned}$$

converges in distribution as stated in the theorem if  $c_x$  is non-zero. This verifies the claims and proves the theorem.  $\square$

### 2.3. Component distributions

To estimate the component distributions let  $A, B$  be two sets that partition  $\mathcal{X}$ . Equations (1.5) and (1.6) then suggest the estimators

$$\begin{aligned} H_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^+(B, A)} (F_n(y|A) - F_n(y|B)), \\ G_n(y; A, B) &\equiv F_n(y|A) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)). \end{aligned} \quad (2.12)$$

Fix  $A, B$  and omit them as arguments, so  $G_n(y) = G_n(y; A, B)$  and  $H_n(y) = H_n(y; A, B)$ .

To state their asymptotic behavior, let

$$\begin{aligned} d_G(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^-(B, A))^2}, \\ d_H(A, B; y) &\equiv \frac{F(y|A) - F(y|B)}{(1 - \zeta^+(B, A))^2}, \end{aligned}$$

and let  $\|\cdot\|_\infty$  denote the supremum norm.

**THEOREM 5.** *Under the conditions of Theorem 2,*

$$\|G_n - G\|_\infty = o_p(1), \quad \|H_n - H\|_\infty = o_p(1),$$

as  $n \uparrow +\infty$ .

*Under the conditions of Theorem 3,*

$$\begin{aligned} \sqrt{\iota_{n_A}}(G_n(y) - G(y)) &\xrightarrow{d} \mathcal{N}(0, d_G(A, B; y)^2 \sigma_-^2(B, A)), \\ \sqrt{\kappa_{n_A}}(H_n(y) - H(y)) &\xrightarrow{d} \mathcal{N}(0, d_H(A, B; y)^2 \sigma_+^2(B, A)), \end{aligned}$$

as  $n \uparrow +\infty$  for each  $y \in \mathbb{R}_+$ .

**Proof.** Consistency follows by Theorem 2 and the Glivenko-Cantelli theorem.

We establish the asymptotic distribution of  $G_n$ ; the result for  $H_n$  follows by the same argument.

First note that

$$\sqrt{\iota_{n_A}}(G_n(y) - G(y)) = T_1 + T_2 + T_3$$

for

$$\begin{aligned} T_1 &\equiv \sqrt{\iota_{n_A}}(F_n(y|A) - F(y|A)), \\ T_2 &\equiv -\frac{1}{1 - \zeta^-(B, A)} \sqrt{\iota_{n_A}} \left( \{F_n(y|A) - F(y|A)\} - \{F_n(y|B) - F(y|B)\} \right), \\ T_3 &\equiv -(F_n(y|A) - F_n(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right). \end{aligned}$$

By the Glivenko-Cantelli theorem,  $T_1 = o_p(1)$  and  $T_2 = o_p(1)$  while

$$T_3 = -(F(y|A) - F(y|B)) \sqrt{\iota_{n_A}} \left( \frac{1}{1 - \zeta_n^-(B, A)} - \frac{1}{1 - \zeta^-(B, A)} \right) + o_p(1).$$

A linearization of this expression in  $\zeta_n^-(B, A) - \zeta^-(B, A)$  together with an application of Theorem 2 to the partition  $A, B$  then yields the result.  $\square$

### 3. SPECIFICATION TESTING

The tail restriction in Assumption 3 and the exclusion restriction on  $X$  are the two key ingredients that lead to identification via Theorem 1. The proof of the theorem shows that both restrictions have testable implications.

#### 3.1. Testing for tail dominance

As is evident from the proof of Theorem 1, Assumption 3 implies that

$$\lambda(x) = \frac{1 - \zeta^-(A, x)}{\zeta^+(A, x) - \zeta^-(A, x)}$$

for any  $A \subset \mathcal{X} - x$ . Denote  $\lambda_n(x; A)$  the corresponding plug-in estimator of  $\lambda(x)$ . Any such estimator will be consistent and asymptotically-normal under the conditions of Theorem 2 and Theorem 3, respectively. Therefore, letting  $A, B$  be two disjoint subsets of  $\mathcal{X} - x$ , a simple specification test for the tail restriction amounts to testing whether the difference

$$\lambda_n(x; A) - \lambda_n(x; B)$$

is statistically different from zero.

As before, the limit distribution of this difference depends on  $c_x = \lim_{n \uparrow +\infty} \iota_{n_x} / \kappa_{n_x}$ .

We define

$$\begin{aligned} \Sigma_- &\equiv d^-(x, A) \left\{ d^-(x, A) \sigma_-^2(A, x) - d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) \right\} \\ &\quad + d^-(x, B) \left\{ d^-(x, B) \sigma_-^2(B, x) - d^-(x, A) \zeta^-(A, x) \zeta^-(B, x) \right\}, \end{aligned}$$

and

$$\begin{aligned} \Sigma_+ \equiv & d^+(x, A) \left\{ d^+(x, A) \sigma_+^2(A, x) - d^+(x, B) \zeta^+(A, x) \zeta^+(B, x) \right\} \\ & + d^+(x, B) \left\{ d^+(x, B) \sigma_+^2(B, x) - d^+(x, A) \zeta^+(A, x) \zeta^+(B, x) \right\}, \end{aligned}$$

in Theorem 6.

THEOREM 6 (A test for tail dominance). *Under the conditions of Theorem 3,*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{\Sigma_- + c_x \Sigma_+ / \sqrt{\ell_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x < +\infty$ , while

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{\lambda_n(x, A) - \lambda_n(x, B)}{\sqrt{c_x^{-1} \Sigma_- + \Sigma_+ / \sqrt{\kappa_{n_x}}}} \right| > z(\tau/2) \right\} = \tau$$

if  $c_x > 0$ , where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** Fix the sets  $A, B$ . We treat only the case  $c_x > 0$ . Our assumptions imply the joint limit result

$$\sqrt{\kappa_{n_x}} \begin{pmatrix} \lambda_n(x; A) - \lambda(x) \\ \lambda_n(x; B) - \lambda(x) \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\kappa^{AA} & \Sigma_\kappa^{AB} \\ \Sigma_\kappa^{AB} & \Sigma_\kappa^{BB} \end{pmatrix} \right),$$

where

$$\begin{aligned} \Sigma_\kappa^{AA} &\equiv c_x^{-1} d^-(x, A)^2 \sigma_-^2(A, x) + d^+(x, A)^2 \sigma_+^2(A, x), \\ \Sigma_\kappa^{BB} &\equiv c_x^{-1} d^-(x, B)^2 \sigma_-^2(B, x) + d^+(x, B)^2 \sigma_+^2(B, x), \end{aligned}$$

follow directly from an application of Theorem 4, and

$$\Sigma_\kappa^{AB} \equiv c_x^{-1} d^-(x, A) d^-(x, B) \zeta^-(A, x) \zeta^-(B, x) + d^+(x, A) d^+(x, B) \zeta^+(A, x) \zeta^+(B, x)$$

follows readily from the asymptotic-linearity statement at the end of the proof of Theorem 3. It follows that the transformation  $\sqrt{\kappa_{n_x}}(\lambda_n(x) - \lambda_n(x))$  is asymptotically normal with variance

$$\Sigma_\kappa^{AA} + \Sigma_\kappa^{BB} - 2 \Sigma_\kappa^{AB} = c_x^{-1} \Sigma_- + \Sigma_+.$$

This proves the result.  $\square$

### 3.2. Testing the exclusion restriction

An implication of the exclusion restriction in (1.1) is that the estimators of the component distributions  $G$  and  $H$  in (2.12), when based on different subsets of  $\mathcal{X}$ , should co-incide with one another, up to sampling error.

Theorem 7 provides the relevant asymptotic variance to perform this test. In it we use

$$\begin{aligned}\Sigma_G = & d_G(A, C; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\} \\ & + d_G(A, B; y) \{d_G(A, C; y) \sigma_-^2(C, A) - d_G(A, B; y) \zeta^-(C, A) \zeta^-(B, A)\}\end{aligned}$$

and

$$\begin{aligned}\Sigma_H = & d_H(A, C; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\} \\ & + d_H(A, B; y) \{d_H(A, C; y) \sigma_+^2(C, A) - d_H(A, B; y) \zeta^+(C, A) \zeta^+(B, A)\},\end{aligned}$$

where the triple  $A, B, C$  constitutes any partition of  $\mathcal{X}$ .

**THEOREM 7** (A test of the exclusion restriction). *Under the conditions of Theorem 3*

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{G_n(y; A, B) - G_n(y; A, C)}{\sqrt{\Sigma_G}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

and

$$\lim_{n \uparrow +\infty} \mathbb{P} \left\{ \left| \frac{H_n(y; A, B) - H_n(y; A, C)}{\sqrt{\Sigma_H}/\sqrt{\iota_{n_A}}} \right| > z(\tau/2) \right\} = \tau,$$

where  $z(\tau)$  is the  $1 - \tau$  quantile of the standard-normal distribution.

**Proof.** We consider only the case of  $G$ . The difference  $G_n(y; A, B) - G_n(y; A, C)$  equals

$$\frac{1}{1 - \zeta_n^-(C, A)} (F_n(y|A) - F_n(y|C)) - \frac{1}{1 - \zeta_n^-(B, A)} (F_n(y|A) - F_n(y|B)).$$

Proceeding as in the proof to Theorem 6 immediately yields the asymptotic-distribution result that implies the theorem.  $\square$

## 4. SIMULATION EXPERIMENTS

In our numerical illustrations we will work with the family of skew-normal distributions (Azzalini 1985.) The skew-normal distribution with location  $\mu$ , positive scale  $\sigma$ , and

skewness parameter  $\beta$  multiplies the density of  $\mathcal{N}(\mu, \sigma^2)$  by a term that skews it to the right if  $\beta > 0$  and to the left if  $\beta < 0$ . Moreover,

$$f(x; \mu, \sigma, \beta) \equiv \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \times \frac{\Phi\left(\beta \frac{x - \mu}{\sigma}\right)}{\Phi(0)}.$$

Its mean and variance are  $\mu + \sigma\delta\sqrt{\frac{2}{\pi}}$  and  $\sigma^2\left(1 - \frac{2\delta^2}{\pi}\right)$ , respectively, where  $\delta \equiv \beta/\sqrt{1 + \beta^2}$ . Clearly,

$$f(x; \mu, \sigma, \beta) \rightarrow \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)$$

as  $\beta \rightarrow 0$ .

In our simulations we will consider data generating processes where the outcome is generated as

$$Y = T V_G + (1 - T) V_H, \quad (4.1)$$

where  $T$  is a latent binary variable, and  $V_G \sim G$  and  $V_H \sim H$ . Both error distributions  $G$  and  $H$  are skewed-normal distributions with parameters  $\mu_G, \sigma_G, \beta_G$  and  $\mu_H, \sigma_H, \beta_H$ , respectively.

From [Capitanio \(2010\)](#) it follows that Assumption 8 holds if  $G$  is right-skewed and  $H$  is left-skewed. We will consider designs where  $\beta_G > 0$  and  $\beta_H < 0$  to verify our asymptotics.

When  $\beta_G = \beta_H = 0$ , (4.1) collapses to a standard location model with normal errors

$$Y = (\mu_G - \mu_H) T + V, \quad V \sim \mathcal{N}(0, \sigma_G^2 + \sigma_H^2). \quad (4.2)$$

The identifying tail condition in Assumption 3 still holds if  $\mu_G > \mu_H$ , and our estimators remain consistent. However, Assumption 8 now fails and so we may expect poor inference in this design.

In our experiments we generate a binary  $X$  with  $\mathbb{P}(X = 1) = \frac{1}{2}$  and fix conditional probabilities as

$$\begin{aligned} \mathbb{P}(T = 0|X = 0) &= \frac{3}{4}, & \mathbb{P}(T = 1|X = 0) &= \frac{1}{4}, \\ \mathbb{P}(T = 1|X = 1) &= \frac{1}{4}, & \mathbb{P}(T = 0|X = 1) &= \frac{3}{4}. \end{aligned}$$

We present results for data generating processes where  $\mu_G = \mu = -\mu_H$  and  $\beta_G = \beta = -\beta_H$ . We use the designs  $\mu = 0$  and  $\beta \in \{2.5, 5\}$  to evaluate the adequacy of our

asymptotic arguments for small-sample inference. We also look at the performance of our estimators when  $\mu \in \{.5, 1\}$  and  $\beta = 0$ , which yields the Gaussian location model in (4.2). We fix  $\sigma_G = \sigma_H = 1$  throughout. For each of these designs we consider choices of the empirical quantiles as

$$\iota_{n_x} = C(n_x \ln \ln n_x)^{6/10}, \quad \kappa_{n_x} = C(n_x \ln \ln n_x)^{6/10},$$

for several choices of the constant  $C$ . All of these choices are in line with our asymptotic arguments. The larger the constant  $C$  the more conservative the choice of intermediate quantile,

$$q_\ell \equiv \frac{\iota_{n_x}}{n_x}, \quad q_r \equiv \frac{n_x - \kappa_{n_x}}{n_x},$$

for a given sample size.

We run experiments for sample sizes  $n \in \{500; 1,000, 2,500; 5,000; 10,000; 25,000\}$ . We report (the average over the replications of)  $q_\ell$  and  $q_r$  along with the estimation results to get an idea of how far in the tails of the component distributions we are going to obtain the results. A data-driven determination of the constant  $C$  is challenging and is left for future research. All tables and figures concerning our Monte Carlo experiment are collected at the end of the paper.

Tables 1–4 report the simulation results for the mixing proportions  $\lambda(0)$  and  $\lambda(1)$  for the data generating processes with  $\mu = 0, \beta = 5$ ;  $\mu = 0, \beta = 2.5$ ;  $\mu = 1, \beta = 0$ ; and  $\mu = .5, \beta = 0$ , respectively. Each table contains the bias, standard deviation (SD), ratio of the (average over the replications of the) estimated standard error to the standard deviation (SE/SD), and the coverage of 95% confidence intervals (CI95). All these statistics were computed from 10,000 Monte Carlo replications.

The results in Table 1 and Table 2 support our asymptotic theory. For each design and all choices of the tuning parameter  $C$ , the bias and standard deviation shrink to zero as  $n \uparrow +\infty$ ; and the bias is small relative to the standard error. Furthermore,  $\text{SE}/\text{SD} \rightarrow 1$  and the coverage rates of the confidence intervals are close to .95 in large samples. The variability of the point estimates is somewhat overestimated when  $n$  is very small and  $C$  is chosen conservatively. Together with the relatively small bias, this implies that confidence intervals are slightly conservative. For  $C = .5$ , coverage rates are close to .95,



even for the smallest samples considered, and for all  $C$ , the coverage rates move fairly quickly toward .95 as  $n$  increases.

Tables 3 and 4 present the results for the pure location model with Gaussian errors, where the tail conditions of Assumption 8 fail. The difference between the two designs is the distance between the component distributions. For Table 3,  $G$  is centered at 1 while  $H$  is centered at  $-1$ , so that  $\mu_G - \mu_H = 2$ . For Table 4,  $G$  and  $H$  are closer to each other:  $\mu_G - \mu_H = 1$ . In the first of these designs the bias in the point estimates is somewhat larger than in the skewed designs. Nonetheless, the bias is still small relative to the standard deviation. Furthermore, the coverage of the confidence intervals displays a similar pattern as in Tables 1 and 2, and is excellent for all values of  $C$  when  $n$  is not too small. When we move to Table 4 the bias increases further. The bias still shrinks to zero as  $n$  grows, confirming that our estimator remains consistent. However, the bias is not negligible relative to the standard deviation; the coverage of the confidence intervals deteriorates as  $n$  grows, and inference becomes unreliable.

We next turn to the results for the component distributions. For clarity we present the results by means of a series of plots. Figures 1–4 contain the results corresponding to Tables 1–4, respectively. To save on space we only present results for one value of  $C$ —the effect of larger values of  $C$  on  $G_n$  and  $H_n$  is similar to that on  $\lambda_n$  documented before. Each figure presents results for  $G_n$  (left plots) and  $H_n$  (right plots) for (from top to bottom)  $n = 1,000$ ,  $n = 2,500$ ,  $n = 5,000$ , and  $n = 10,000$ . Each plot contains the mean of the point estimates and the mean of 95% confidence bounds constructed using a plug-in estimator of the asymptotic variance in Theorem 5 (dashed lines). Each plot also contains the true component distribution and 95% confidence bounds constructed using the empirical standard deviation over the Monte Carlo replications (solid grey lines). We vary the range of the vertical axis across the plots in a given figure to enhance visibility.

Figures 1 and 2 again confirm our asymptotics. Indeed, moving down the rows,  $G_n$  and  $H_n$  converge to  $G$  and  $H$  respectively, while the estimated standard error converges to the Monte Carlo standard deviation. The bias in the point estimates is small across all plots. For small  $n$  the estimated standard errors are somewhat too small; this underestimation is more severe for  $H_n$  than for  $G_n$ .

Figures 3 and 4 concern the Gaussian location model. The results are in line with

our findings concerning the mixing proportions. In the design where  $\mu_G - \mu_H = 2$  our estimators do well in spite of Assumption 8 not holding. In Figure 4, however, the asymptotic bias in  $G_n$  and  $H_n$  becomes visible. While the variability of the point estimates is correctly captured by our asymptotic-variance estimator, the confidence bounds settle around an incorrect curve.

### CONCLUDING REMARKS

We conducted most of our analysis with a mixture of two components. However, some of our results would extend to a version of (1.1) with a larger number of components. Suppose that the mixture has  $J$  irreducible components, as in

$$F(y|x) = \sum_{j=1}^J \lambda_j(x) G_j(y),$$

in obvious notation. Henry et al. (2014) showed that the mixture components and mixing proportions are only identified up to  $J(J-1)$  inequality-constrained real parameters in general.

Tail dominance restrictions can still be quite powerful. Take  $J = 3$  for instance, and assume that  $G_1$  dominates in the left tail and  $G_3$  dominates in the right tail. Then it is easy to adapt the proof of Theorem 1 to prove that the behavior of  $F(y|x)$  in the left tail identifies the function  $\lambda_1$  up to a multiplicative constant, and that the behavior of  $F(y|x)$  in the right tail identifies the function  $\lambda_3$  up to another multiplicative constant. Imposing the values of the mixing proportions at one particular value of  $x$  would be enough to point identify all elements of the model, for instance; and it would be easy to adapt our estimators and tests to such a setting. Whether such additional restrictions are plausible is, of course, highly model-dependent.

**Table 1.**  $\mu = 0$  and  $\beta = 5$ 

$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0115	−.0083	.0951	.0914	1.0790	1.0384	.9714	.9694
1,000	.059	.940	.0060	−.0059	.0693	.0701	1.0554	1.0392	.9688	.9682
2,500	.043	.957	.0038	−.0029	.0510	.0506	1.0321	1.0393	.9624	.9636
5,000	.034	.966	.0016	−.0023	.0406	.0407	1.0246	1.0246	.9588	.9592
10,000	.026	.974	.0012	−.0011	.0328	.0325	1.0106	1.0213	.9560	.9572
25,000	.018	.982	.0013	−.0007	.0245	.0244	1.0108	1.0149	.9526	.9568
$C = 1$										
500	.172	.848	.0044	−.0046	.0534	.0535	1.1937	1.1907	.9904	.9876
1,000	.120	.880	.0024	−.0035	.0439	.0446	1.1358	1.1220	.9764	.9752
2,500	.087	.914	.0012	−.0010	.0332	.0337	1.1031	1.0847	.9680	.9668
5,000	.067	.933	.0012	−.0005	.0270	.0271	1.0814	1.0726	.9664	.9670
10,000	.052	.947	.0007	−.0003	.0225	.0222	1.0360	1.0519	.9566	.9616
25,000	.037	.963	.0006	−.0005	.0167	.0167	1.0459	1.0460	.9594	.9600
$C = 1.5$										
500	.227	.773	.0198	−.0198	.0328	.0333	1.6715	1.6482	.9982	.9984
1,000	.179	.821	.0046	−.0037	.0316	.0315	1.2931	1.2933	.9944	.9920
2,500	.129	.870	.0013	.0000	.0257	.0256	1.1575	1.1579	.9774	.9762
5,000	.101	.898	.0010	−.0005	.0208	.0215	1.1427	1.1070	.9756	.9720
10,000	.078	.922	.0002	−.0010	.0175	.0174	1.0873	1.0962	.9646	.9710
25,000	.056	.944	.0004	.0000	.0136	.0136	1.0453	1.0474	.9632	.9604

**Table 2.**  $\mu = 0$  and  $\beta = 2.5$ 

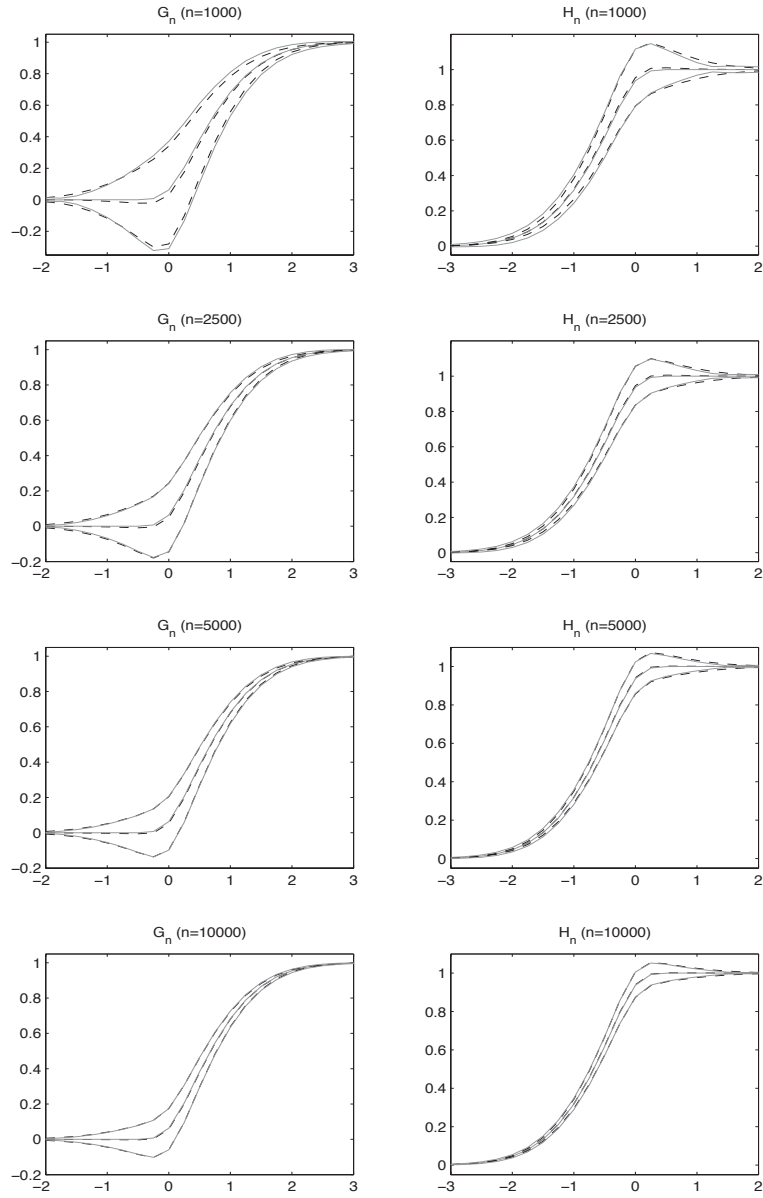
	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0093	−.0122	.0941	.0934	1.0161	1.0315	.9718	.9780
1,000	.059	.940	.0066	−.0072	.0722	.0718	1.0151	1.0194	.9646	.9652
2,500	.043	.957	.0029	−.0040	.0512	.0519	1.0279	1.0212	.9610	.9612
5,000	.034	.966	.0016	−.0026	.0407	.0407	1.0248	1.0255	.9546	.9596
10,000	.026	.974	.0012	−.0015	.0323	.0326	1.0287	1.0193	.9548	.9626
25,000	.018	.982	.0007	−.0009	.0243	.0244	1.0182	1.0137	.9564	.9580
$C = 1$										
500	.172	.848	.0160	−.0174	.0519	.0518	1.2924	1.2969	.9954	.9972
1,000	.120	.880	.0069	−.0084	.0429	.0430	1.1865	1.1898	.9862	.9890
2,500	.087	.914	.0032	−.0024	.0334	.0338	1.1035	1.0900	.9716	.9684
5,000	.067	.933	.0018	−.0014	.0270	.0275	1.0868	1.0648	.9700	.9666
10,000	.052	.947	.0001	−.0010	.0220	.0219	1.0569	1.0661	.9636	.9652
25,000	.037	.963	.0005	−.0005	.0168	.0168	1.0376	1.0356	.9603	.9585
$C = 1.5$										
500	.227	.773	.0491	−.0496	.0361	.0365	1.7108	1.6925	.9908	.9902
1,000	.179	.821	.0245	−.0248	.0304	.0303	1.4562	1.4670	.9920	.9926
2,500	.129	.870	.0083	−.0080	.0248	.0250	1.2350	1.2249	.9868	.9872
5,000	.101	.898	.0030	−.0035	.0211	.0210	1.1386	1.1418	.9774	.9788
10,000	.078	.922	.0015	−.0016	.0172	.0172	1.1103	1.1090	.9722	.9720
25,000	.056	.944	.0005	−.0007	.0136	.0136	1.0496	1.0455	.9604	.9568

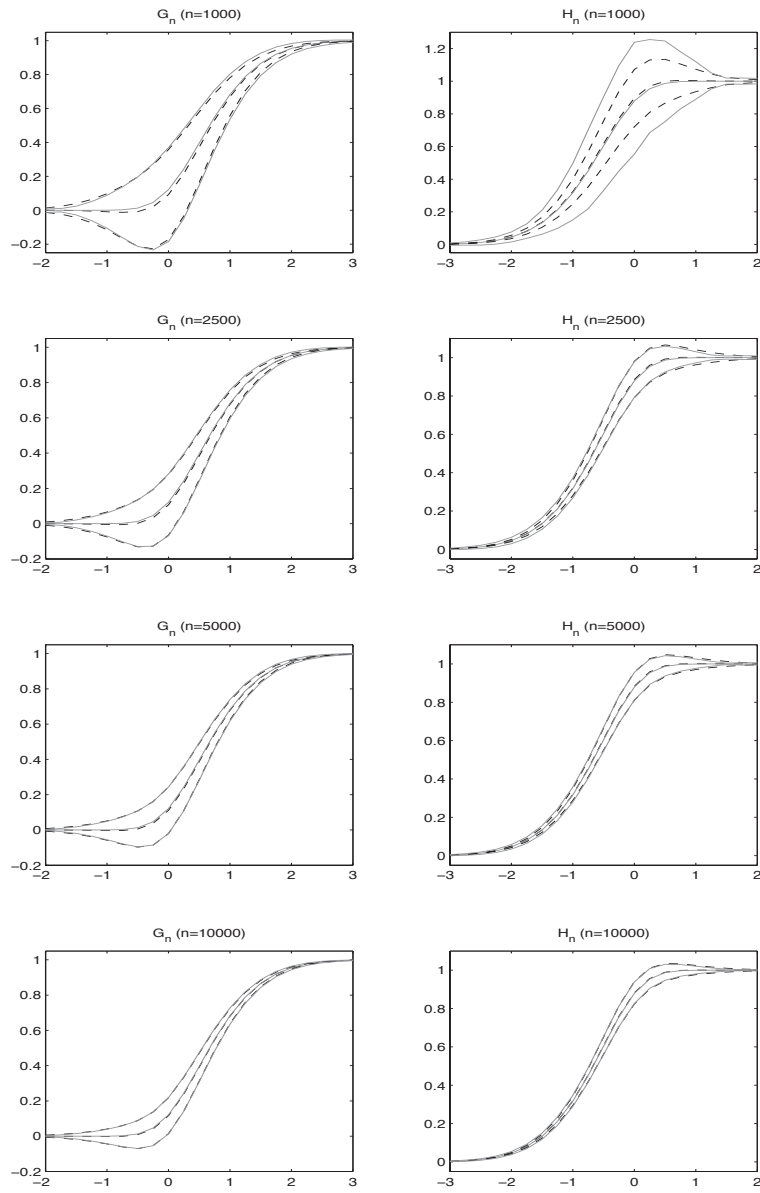
**Table 3.**  $\mu = 1$  and  $\beta = 0$ 

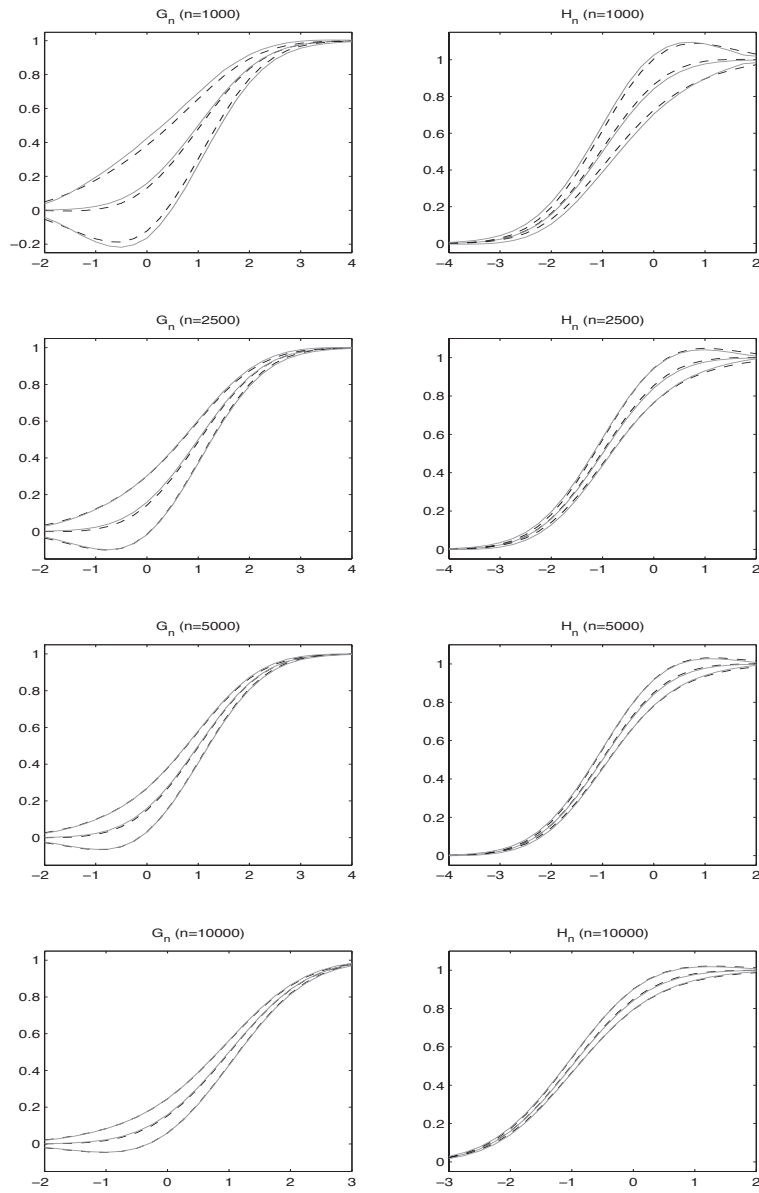
$n$	$q_\ell$	$q_r$	BIAS		SD		SE/SD		CI95	
			$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .5$										
500	.075	.924	.0220	−.0227	.0966	.0948	1.0479	1.0710	.9869	.9866
1,000	.059	.940	.0144	−.0164	.0720	.0728	1.0589	1.0518	.9807	.9810
2,500	.043	.957	.0092	−.0105	.0519	.0527	1.0476	1.0331	.9705	.9697
5,000	.034	.966	.0067	−.0069	.0411	.0415	1.0372	1.0283	.9673	.9651
10,000	.026	.974	.0050	−.0048	.0327	.0324	1.0344	1.0449	.9614	.9622
25,000	.018	.982	.0030	−.0033	.0246	.0245	1.0188	1.0246	.9576	.9586
$C = 1$										
500	.172	.848	.0385	−.0395	.0576	.0575	1.2942	1.2958	.9948	.9944
1,000	.120	.880	.0268	−.0267	.0459	.0453	1.2162	1.2304	.9867	.9869
2,500	.087	.914	.0154	−.0166	.0339	.0344	1.1554	1.1417	.9786	.9770
5,000	.067	.933	.0110	−.0120	.0274	.0278	1.1168	1.1036	.9722	.9691
10,000	.052	.947	.0079	−.0084	.0225	.0224	1.0752	1.0823	.9641	.9656
25,000	.037	.963	.0055	−.0055	.0169	.0173	1.0602	1.0361	.9603	.9562
$C = 1.5$										
500	.227	.773	.0718	−.0714	.0421	.0418	1.6251	1.6382	.9726	.9739
1,000	.179	.821	.0476	−.0474	.0343	.0340	1.4396	1.4479	.9619	.9632
2,500	.129	.870	.0277	−.0282	.0269	.0270	1.2510	1.2445	.9502	.9487
5,000	.101	.898	.0188	−.0191	.0217	.0219	1.1929	1.1831	.9510	.9473
10,000	.078	.922	.0133	−.0132	.0178	.0179	1.1330	1.1285	.9443	.9469
25,000	.056	.944	.0086	−.0084	.0135	.0136	1.0949	1.0895	.9424	.9458

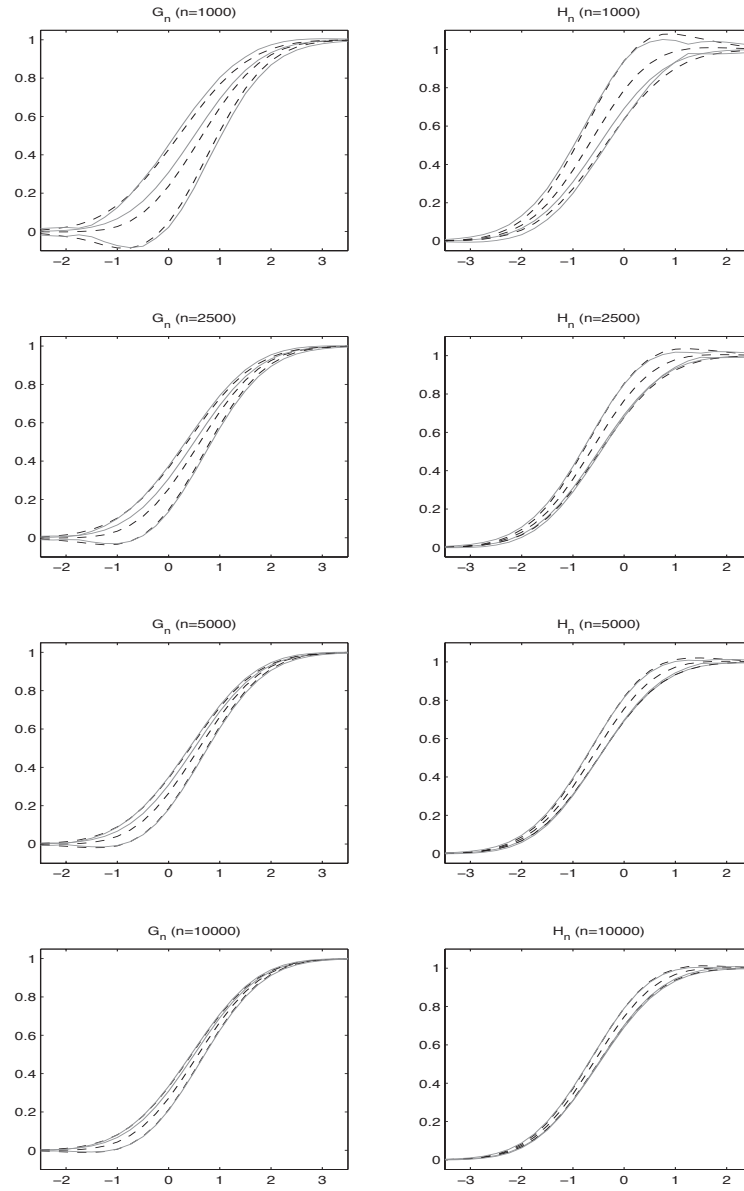
**Table 4.**  $\mu = .5$  and  $\beta = 0$ 

	BIAS				SD		SE/SD		CI95	
$n$	$q_\ell$	$q_r$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$	$\lambda_n(0)$	$\lambda_n(1)$
$C = .75$										
500	.114	.886	.1131	−.1145	.1162	.1147	1.1992	1.2238	.9864	.9880
1,000	.090	.910	.0994	−.1017	.0842	.0855	1.1677	1.1599	.9416	.9406
2,500	.065	.935	.0840	−.0864	.0594	.0607	1.1175	1.0999	.8470	.8404
5,000	.051	.949	.0755	−.0758	.0454	.0463	1.1092	1.0920	.7448	.7342
10,000	.039	.961	.0671	−.0671	.0358	.0352	1.0815	1.0973	.6244	.6286
25,000	.028	.972	.0580	−.0573	.0257	.0259	1.0710	1.0601	.4374	.4584
$C = 1.25$										
500	.189	.811	.1451	−.1445	.0905	.0915	1.3579	1.3417	.9196	.9150
1,000	.149	.851	.1258	−.1271	.0673	.0682	1.2780	1.2632	.7896	.7808
2,500	.108	.892	.1075	−.1065	.0476	.0475	1.2051	1.2037	.5524	.5472
5,000	.084	.916	.0938	−.0947	.0369	.0371	1.1590	1.1555	.3806	.3672
10,000	.065	.935	.0832	−.0837	.0284	.0284	1.1426	1.1407	.2146	.2208
25,000	.047	.953	.0713	−.0714	.0205	.0208	1.1143	1.1002	.0740	.0740
$C = 1.75$										
500	.265	.735	.1704	−.1723	.0747	.0757	1.5656	1.5429	.8094	.7980
1,000	.209	.791	.1498	−.1500	.0573	.0580	1.4223	1.4072	.5592	.5578
2,500	.152	.849	.1255	−.1253	.0408	.0417	1.2970	1.2674	.2658	.2702
5,000	.118	.882	.1114	−.1095	.0323	.0318	1.2143	1.2267	.1076	.1160
10,000	.092	.908	.0968	−.0980	.0247	.0248	1.1889	1.1849	.0358	.0296
25,000	.065	.935	.0827	−.0827	.0178	.0180	1.1481	1.1324	.0034	.0030

**Figure 1.**  $\mu = 0$  and  $\beta = 5$  ( $C = .5$ )

**Figure 2.**  $\mu = 0$  and  $\beta = 2.5$  ( $C = .5$ )

**Figure 3.**  $\mu = 1$  and  $\beta = 0$  ( $C = .5$ )

**Figure 4.**  $\mu = .5$  and  $\beta = 0$  ( $C = .75$ )



## REFERENCES

- Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tabaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica* 80(5), 1977–2016.
- Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* 37, 3099–3132.
- Andrews, D. W. K. and M. M. A. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.
- Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review* 102, 94–130.
- Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature* 49, 3–71.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.
- Bollinger, C. R. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics* 73, 387–399.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014a). Estimating multivariate latent-structure models. Sciences Po Discussion Paper 2014-18.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2014b). Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society - Series B*, forthcoming.
- Bordes, L., S. Mottelet, and P. Vandekerckhove (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics* 34, 1204–1232.
- Capitanio, A. (2010). On the approximation of the tail probability of the scalar skew-normal distribution. *METRON* 68, 299–308.
- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall, CRC Press.
- D’Haultfoeuille, X. and P. Février (2014). Identification of mixture models using support variations. Mimeo.
- D’Haultfoeuille, X. and A. Maurel (2013). Another look at identification at infinity of sample selection models. *Econometric Theory* 29, 213–224.
- Einmahl, J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *Journal of Statistical Planning and Inference* 32, 137–145.
- Einmahl, U. and D. Mason (1997). Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields* 107, 283–311.
- Frisch, R. (1934). Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics* 1, 255–294.
- Gassiat, E. and J. Rousseau (2014). Non parametric finite translation mixtures with dependent regime. *Bernoulli*, forthcoming.
- Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics Volume 14: Statistical Methods in Finance*. Elsevier.

- Hall, P. and X.-H. Zhou (2003). Nonparametric identification of component distributions in a multivariate mixture. *Annals of Statistics* 31, 201–224.
- Hamilton, J. D. (1989). A new approach to the analysis of nonstationary times series and the business cycle. *Econometrica* 57, 357–384.
- Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.
- Heckman, J. J. (1990). Varieties of selection bias. *American Economic Review* 80, 313–318.
- Henry, M., Y. Kitamura, and B. Salanié (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Paper 1767.
- Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5, 123–144.
- Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.
- Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77, 135–175.
- Khan, S. and E. Tamer (2010). Irregular identification, support conditions and inverse weight estimation. *Econometrica* 78, 2021–2042.
- Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.
- Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica* 74, 631–665.
- Schwarz, M. and S. Van Belleghem (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters* 80, 236–241.
- Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica* 68, 343–369.